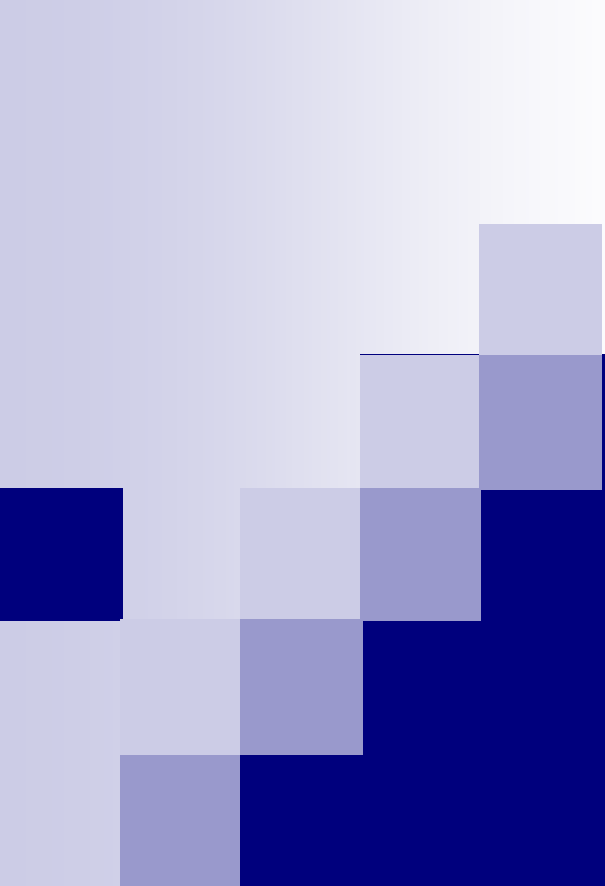




Statistique descriptive

sigmakutub.blogspot.com

Licence fondamentale Economie Gestion



sigmakutub.com

Bibliographie

- 
- TENENHAUS Michel, « Statistique : Méthodes pour décrire, expliquer et prévoir », DUNOD, 2006.
 - LETHIELLEUX Maurice, « Statistique descriptive », DUNOD, 2003.
 - CHAUVAT Gérard, REAU Jean-Philippe, « Statistiques descriptives », ARMAND COLIN, 2002.
 - MONINO Jean-Louis, KOSIANSKI Jean-Michel, LE CORNU François, « Statistiques descriptives - Travaux dirigés », DUNOD, 2000.
 - GOLDFARB Bernard, PARDOUX Catherine, « Introduction à la Méthode Statistique », DUNOD, 2003.
 - GOLDFARB Bernard, PARDOUX Catherine, « Introduction à la Méthode Statistique, Exercices Corrigés », DUNOD, 2003.
 - PY Bernard, « Statistique Descriptive », ECONOMICA, Dernière édition.
 - PY Bernard, « Exercices corrigés de Statistique Descriptive », ECONOMICA.

Plan du chapitre introductif

- 
- I. Définition de la statistique
 - II. Approches de la statistique
 - a. *Statistique descriptive*
 - b. *Statistique mathématique*
 - III. Domaines d'application de la statistique
 - IV. Vocabulaire statistique : Définitions
 - V. Types de critères ou de variables
 - a. Variables quantitatives
 - i. Variables quantitatives discrètes
 - ii. Variables quantitatives continues
 - b. Variables qualitatives
 - i. Variables qualitatives ordinales
 - ii. Variables qualitatives nominales

Chapitre introductif (1)

I. Définition de la statistique

« La statistique a pour objet l'étude, à l'aide de traitements mathématiques, de nombreux faits correspondant à l'observation d'un phénomène, dans le but de rendre compte de la réalité, d'essayer de l'expliquer et d'aider à la prise de décision » (J. Hubler, 1996)

➤ **Ex** : Le RGP est un gisement de connaissances sur les aspects démographiques, économiques et sociaux de la population marocaine à un moment donné

➔ Ses résultats doivent permettre d'orienter les décisions de l'Etat en recherchant des solutions aux problèmes de la pauvreté, de l'habitat insalubre, etc.

■ Ne pas confondre « ***La statistique*** » et « ***Les statistiques*** »

➔ ***La*** statistique : Cf. définition

➔ ***Les*** statistiques : données chiffrées ou les résultats numériques de ***la*** statistique

Chapitre introductif (2)

II. La collecte des données statistiques

- Deux principales sources de données statistiques
 - ➔ Les recensements
 - ➔ Les enquêtes

1. Les recensements

- ➔ Sont des opérations, issues du dénombrement, qui consistent à étudier de façon **exhaustive** et en fonction de plusieurs critères **tous** les éléments d'une population
- ➔ Ne pas confondre « ***dénombrement*** » et « ***recensement*** »
 - ⇒ Le ***dénombrement*** : comptage des individus d'une population
 - ⇒ Le ***recensement*** : chiffrer les données selon plusieurs aspects (âge, sexe, chiffre d'affaires, etc.)

Chapitre introductif (3)

Exemple :

Dénombrement

La population légale du Maroc a atteint 29.891.708 habitants. Selon les premiers résultats du recensement général de la population et de l'habitat, réalisé entre les 2 et 20 septembre 2004, révélés par le Haut commissariat, cette population est répartie en 16.463.634 **citadins** et 13.428.074 **ruraux**, soit un taux d'urbanisation de 55,1 %. **La répartition régionale** est marquée par la concentration de près du tiers de la population dans trois régions : le Grand Casablanca avec ses 3,6 millions d'habitants (12,1%), Souss-Massa-Draâ et Marrakech-Tensift -Al Haouz avec 3,1 millions chacune (10,4 %). Le reste est réparti selon des proportions variant entre 0,3 % (Oued Eddahab-Lagouira) et 8,3 % (Tanger-Tétouan).

.....→ **Données sur la répartition de la pop. : Recensement**

Chapitre introductif (4)

2. Les enquêtes

- Portent sur un sous-ensemble d'une population appelé *échantillon*
- Ne sont pas exhaustives : n'interrogent pas tous les éléments d'une population
- ↪ *La qualité de l'enquête et donc des résultats dépend du choix de l'échantillon*

Chapitre introductif (5)

III. Approches de la statistique

- Statistique descriptive ou déductive
- Statistique mathématique ou inductive

1. **Statistique descriptive** : classification des données et leur traitement afin de les rendre utilisables et permettre leur interprétation

⇒ **Ex** : moyenne générale d'un groupe de TD

2. **Statistique mathématique** : ensemble de méthodes mathématiques qui permettent de faire des **prévisions**, des **interpollations** sur une population à partir de résultats recueillis sur un échantillon

↳ *Raisonnements inductifs* : passage du particulier au général

Chapitre introductif (6)

IV. Domaine d'application de la statistique

- Démographie
- Sciences économiques et sociales
- Sociologie
- Marketing
- Géophysique
- Physique
- Médecine
- Sciences politiques
- Etc.

Chapitre introductif (7)

V. Vocabulaire statistique : Définitions

- **Population** : ensemble des unités statistiques ou individus sur lesquels on effectue une analyse statistique
 - ⇒ Ex : étudiants de Aïn Sebaa; production de voitures Renault en 2000
- **Unités statistiques** (individus): élément de la population sur lequel porte l'observation
 - ⇒ Ex : étudiants; voitures Renault
- **Echantillon** : ensemble d'individus prélevés dans une population déterminée
 - ⇒ Ex : étudiants de moins de 20 ans; Renault Mégane
- **Caractère** (critère): permet de décrire et de classer la population
 - ⇒ Ex : classification des étudiants selon « l'âge » ou « le sexe »; « couleur » ou « puissance » des Megane,

Chapitre introductif (8)

Population urbaine marocaine par groupe d'âge et sexe (en millier)

Groupe d'âge	2005			2006		
	F	M	<i>Total</i>	F	M	<i>Total</i>
0-9	1463	1513	<i>2976</i>	1466	1516	<i>2982</i>
10-30	3265	3210	<i>6475</i>	3298	3244	<i>6542</i>
30-60	3067	2960	<i>6027</i>	3187	3059	<i>6246</i>
60 et +	687	591	<i>1278</i>	706	604	<i>1310</i>
Total	8481	8274	<i>16755</i>	8656	8423	<i>17079</i>

Source : http://doc.abhatoo.net.ma/doc/IMG/pdf/PopduMarocpargroupee_ageetsexede2005-2007.pdf

- **Population** : population urbaine marocaine en 2005 et 2006
- **Individu** : population urbaine
- **Caractère** : groupe d'âge et sexe

Chapitre introductif (9)

VI. Types de critères, de caractères ou de variables

- Caractères quantitatifs
- Caractères qualitatifs

1. Les variables quantitatives

- Variables numériques et mesurables exprimant une quantité
 - ⇒ **Ex** : Chiffre d'Affaires d'une entreprise; taux de chômage; taille; PIB, etc
- Les variables quantitatives peuvent être classées en :
 - Variables quantitatives **discrètes** ou **discontinues***
 - Variables quantitatives **continues***

Chapitre introductif (10)

a. Variable quantitative discrète (discontinue)

- Elle est représentée par un nombre *fini* de valeurs (Ex : nombre d'enfant par ménage; nombre d'hospitalisation par patient, etc.)
- Les modalités de la variable peuvent être traitées mathématiquement (par des opérations mathématiques de base)

b. Variable quantitative continue

- Elle peut prendre un nombre *infini* de valeurs dans son intervalle de définition (Ex : taille, revenus, CA, poids, etc.)
- Il s'agit de grandeurs liées à *l'espace* (longueur, surface), au *temps* (âge, durée, vitesse), à la *masse* (poids, teneur), à la *monnaie* (salaire, CA)
- Les variables continues peuvent être regroupées en *classe* : un individu qui pèse 76,5 Kg sera repéré dans une classe de poids de [76-77]

Chapitre introductif (11)

- Exemple : enquête réalisée auprès de 20 femmes casablancaises nées en 1970 sur le nombre d'enfants qu'elles ont eus

Nombre d'enfants/femmes

<i>Nombre d'enfants</i>	<i>Effectif de femmes</i>
0	1
1	3
2	5
3	5
4	4
5	2
<i>Total</i>	<i>20</i>

Chapitre introductif (12)

- On peut choisir de regrouper les différentes valeurs (modalités) de la variable « enfant » en **classes**

Nombre d'enfants/femmes

<i>Nombre d'enfants</i>	<i>Effectif de femmes</i>
[0-2[	4
[2-4[	10
[4-6[	6
<i>Total</i>	<i>20</i>

Chapitre introductif (13)

- Lorsque les données sont regroupées en classe, il faut définir les **extrémités de classe**
 - ➔ Il faut préciser la « *borne inférieure* » et la « *borne supérieure* » des classes
 - ➔ Il faut préciser sans ambiguïté si *les valeurs des extrémités* sont *inclues* ou *non* dans les classes
- **Exemple 1** : nombre d'enfants par femme
 - **Classe [2 – 4[**
 - ➔ « [2 – » signifie que la valeur « 2 » est incluse dans la classe
 - ➔ « – 4 [» signifie que la valeur « 4 » est exclue de la classe
 - Tous les éléments de la population étudiée (femmes) doivent se retrouver dans une et une seule classe

Chapitre introductif (14)

- **Exemple 2** : Salaires mensuels des employés d'une entreprise « X » en DH au 31/12/2006

→ 3 classes de salaires :

⇒ **De 6000 à moins de 7000 DH:** $[6000 - 7000[$

➤ Cette classe comprendra un employé dont le salaire = 6999 tandis qu'un salarié dont le revenu = 7000 s'en trouvera exclu

⇒ **De 7000 à moins de 9000 DH:** $[7000 - 9000[$

⇒ **De 9000 à moins de 12 000 DH:** $[9000 - 12\ 000[$

→ Pour des raisons pratiques, on retient généralement comme extrémités de classes des valeurs « *rondes* »

⇒ Effectuer aisément des calculs sur les extrémités de classes comme pour le calcul de *l'amplitude des classes* et du *centre des classes*

Chapitre introductif (11)

- **L'amplitude de classe** = la différence entre la valeur de l'extrémité supérieure et la valeur de l'extrémité inférieure

→ L'amplitude a d'une classe i sera donnée par la formule suivante :

$$a_i = e_i^{sup} - e_i^{inf}$$

- **Exemple 1** : L'amplitude a_i de la classe [6000 – 7000[

$$a_i = e_i^{sup} - e_i^{inf} = 7000 - 6000 = 1000$$

- **Exemple 2** : Nombre d'enfants par femme

Nombre d'enfants	Effectifs	Amplitudes a_i
[0 – 2 [	4	2
[2 – 4 [	10	2
[4 – 6 [	6	2

← Les classes sont d'amplitudes égales

Chapitre introductif (12)

- **Exemple 3** : Salaires des employés de l'entreprise « X » en DH

Salaires	Amplitudes a_i
[6000 – 7000[	1000
[7000 – 9000[	2000
[9000 – 12 000[	3000

← Les classes sont d'amplitudes inégales

- ➔ *L'amplitude de la deuxième classe est 2 fois plus grande que celle de la première classe*
- ➔ *L'amplitude de la troisième classe est 3 fois plus grande que celle de la première classe*

Chapitre introductif (13)

- **Le centre de classe** = la moyenne des extrémités de classe

→ Le centre c_i d'une classe i sera donnée par la formule suivante :

$$c_i = \frac{e_i^{sup} + e_i^{inf}}{2}$$

- **Exemple 1** : Cas où les amplitudes sont égales (Nombre d'enfants par femme)

Nombre d'enfants	Amplitudes	Centres c_i
[0 – 2 [	2	1
[2 – 4 [	2	3
[4 – 6 [	2	5

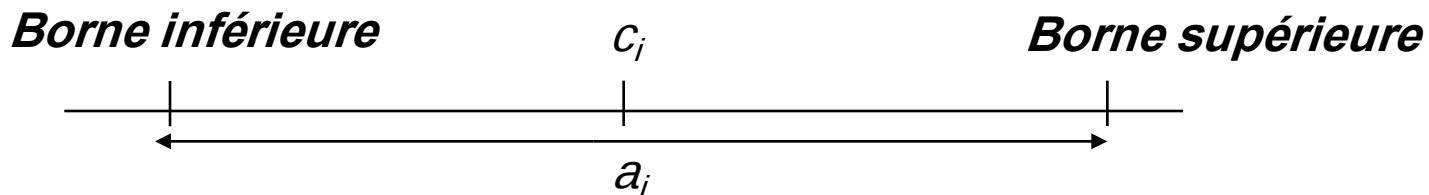
Chapitre introductif (14)

- **Exemple 2** : Cas de classes d'amplitudes inégales (Salaires des employés de l'entreprise « X » en DH)

Salaires	Amplitudes	Centres c_i
[6000 – 7000[	1000	6500
[7000 – 9000[	2000	8000
[9000 – 12 000[	3000	10 500

↪ **Chaque classe est caractérisée par :**

- Borne inférieure
- Borne supérieure
- Amplitude (a_i)
- Centre (c_i)



Chapitre introductif (15)

- **Application** : Répartition des Salaires des employés de l'entreprise « Y » en DH

Salaires	Effectifs	Amplitudes	Centres c_i
[6000 – 7000[	10	1000	6500
[7000 – 9000[	50	2000	8000
[9000 – 10 000[	200	1000	9500
[10 000 – 13 000[	20	3000	11 500
[13 000 – 17 000[	10	4000	15 000
[17 000 – 30 000[	5	13 000	23 500
Total	295	-	-

- ➔ Les classes sont d'amplitudes inégales
- ➔ La troisième classe est mal choisie car l'effectif correspondant est très important par rapport aux autres classes : on aurait pu choisir de la diviser en 2 classes d'amplitudes égales à 500 pour faire apparaître plus d'informations

Chapitre introductif (16)

- 2. Caractère qualitatif : ne peut faire l'objet d'une mesure car il ne se présente pas sous forme numérique. (Ex : couleur de peau; section du bac; catégorie socio-professionnelle; etc.)
 - On ne peut pas effectuer d'opérations arithmétiques sur les caractères qualitatifs (on ne peut additionner les couleurs de peau des êtres humains)
- Les caractères qualitatifs se déclinent en plusieurs **modalités**
 - **Modalités** : les différentes valeurs prises par un caractère qualitatif
 - **Exemple 1** : la variable « **sexe** » à deux modalités « **Masculin** » « **Féminin** »
 - **Exemple 2** : la variable « **couleurs des yeux** » peut prendre comme modalités « **Noir** » « **Brun** » « **Bleu** » « **Vert** » « **Gris** »

Chapitre introductif (17)

→ **Exemple 3** : si la population est décrite selon le caractère « **CSP agrégées** », les différentes modalités seront

Catégories Socio-Professionnelles (CSP) Agrégées
Agriculteurs, exploitants
Artisans, commerçants et chef d'entreprises
Cadres et professions intellectuelles supérieures
Professions intermédiaires
Employés
Ouvriers
Retraités
Autres personnes sans activité professionnelle

← « **Caractère** »

← **Modalités**

Chapitre introductif (18)

- Les modalités d'un caractère qualitatif sont ***exhaustives*** et mutuellement ***incompatibles***

→ **Exhaustives** : à chaque individu doit correspondre une modalité du caractère

Ex : enquête sur l'état matrimonial d'un groupe d'individu

- Pour satisfaire la condition ***d'exhaustivité***, on doit avoir quatre modalités du caractère « Etat matrimonial » : Célibataire, Marié, Veuf, Divorcé

→ **Incompatibles** : Chaque individu doit pouvoir être classé dans **une seule** modalité du caractère

Ex : Un individu ne peut être à la fois « célibataire » et « marié »

↳ *Chaque individu d'un caractère doit pouvoir être classé dans **une** et **une seule** modalité*

Chapitre introductif (19)

- Les modalités d'un caractère qualitatif peuvent être **ordinales** ou **nominales**

→ **Les modalités ordinales** : peuvent être classées ou hiérarchisées

➤ **Ex** : Enquête réalisée en 2006 par l'association « Maroc Entrepreneur » sur le degré de satisfaction des marocains ayant vécu à l'étranger et franchi le cap du retour au Maroc

⇒ **Le Caractère** : « Degré de satisfaction »

⇒ **Les modalités du caractère** : « Satisfait », « Assez Satisfait », « Peu Satisfait », « Pas Satisfait »

Chapitre introductif (20)



Modalités	%
Satisfait	28,06%
Assez Satisfait	33,73%
Peu Satisfait	23,28%
Pas Satisfait	14,93%

- Les modalités sont **ordinales** car on peut les *classer*
 - Le classement effectué va de l'opinion « Satisfait » à l'opinion « Pas Satisfait »
 - On passe d'une préférence *positive* à une préférence de plus en plus *négative*
- Les modalités **ordinales** ne peuvent faire l'objet d'aucune opération arithmétique

Chapitre introductif (21)

→ **Les modalités nominales** : ne peuvent pas être classées (hiérarchisées)

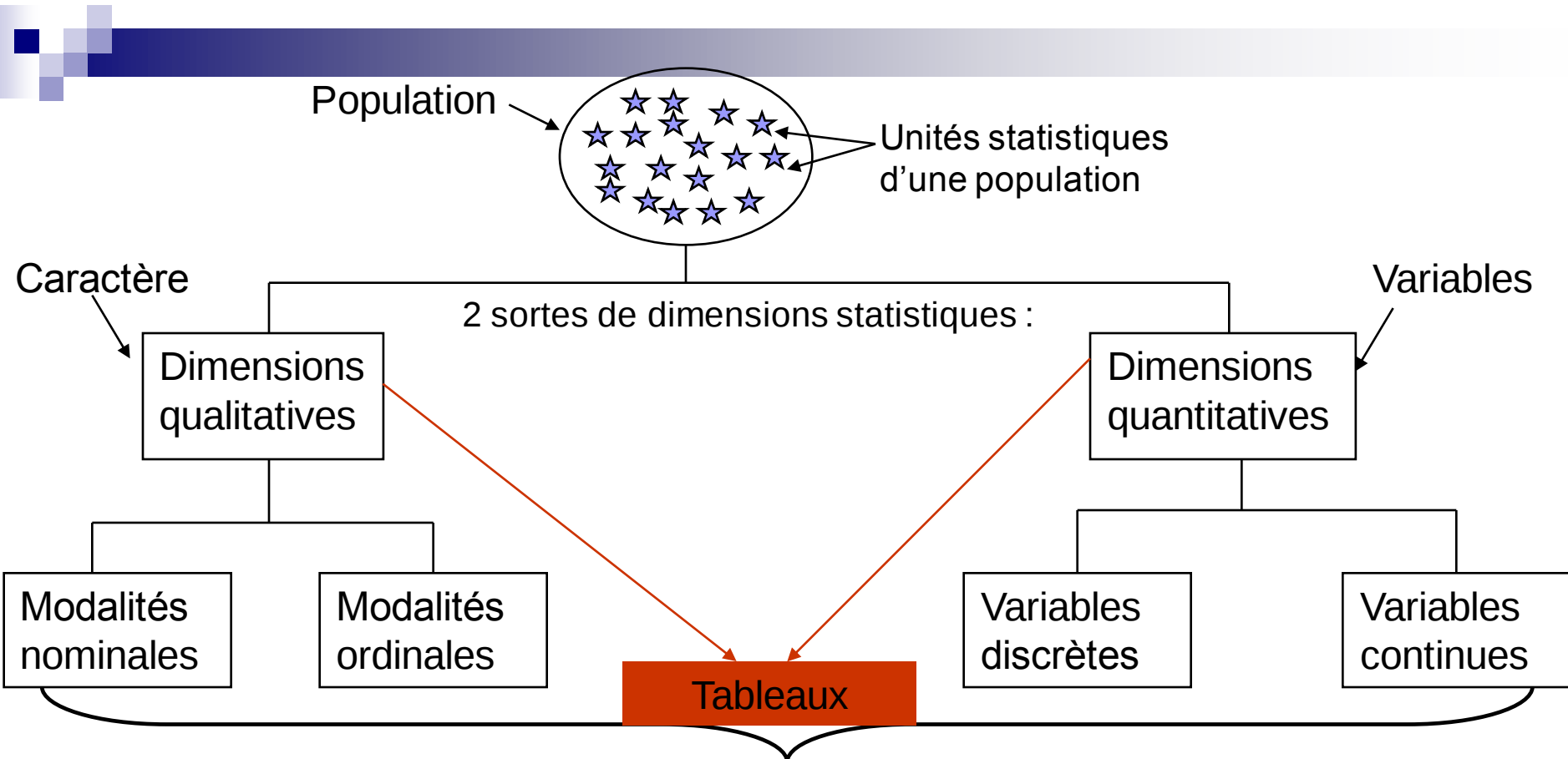
➤ **Ex** : Classement d'un groupe de 15 étudiants selon leur ville de naissance

Modalités	Effectif
Casablanca	9
Mohammedia	4
Rabat	1
El Jadida	1

→ Les 4 modalités du caractère « Ville de naissance » sont **nominales**

⇒ *Les 4 modalités ne peuvent faire l'objet d'aucun classement hiérarchique*

Chapitre introductif (22)



- Données individuelles ou exhaustives
- Données regroupées par valeurs ou par modalités
- Données regroupées par classes de valeurs ou par classes de modalités

Chapitre II – les tableaux statistiques

Plan du chapitre

Introduction

Section 1 : Les tableaux statistiques à un caractère

I. Les tableaux des caractères qualitatifs

- a. Cas de caractère à modalités nominales
- b. Cas de caractère à modalités ordinales

II. Les tableaux des caractères quantitatifs

- a. Cas de caractère quantitatif discret
- b. Cas de caractère quantitatif continue

Chapitre II – les tableaux statistiques

Plan du chapitre

Section 2 : Les tableaux statistiques à deux caractères

- I. Présentation générale des tableaux de contingence
- II. Propriétés structurelles des tableaux de contingence
- III. Les différentes distributions statistiques
- IV. Les relations entre les caractères

Conclusion

Chapitre II – les tableaux statistiques (1)

- L'un des objectifs de la statistique descriptive est de résumer les données « brutes » recueillies sur une population dans des tableaux statistiques

→ Avantages :

- Présentation des données de façon lisible
- En **ligne** : informations relatives à chaque individu
- En **colonne** : critères ou caractères étudiés

- **Exemple 1** : Enquête d'opinion réalisée auprès de 9 étudiants de premières années sciences économiques

→ Données recueillies : Nom, Prénom, Age, série du bac, opinion sur l'architecture de la fac de Aïn Sebaa

- ## → Matrice des données : = {{"Alaoui", "Fatima", 18, "L", "Très bonne"}, {"Otmani", "Samira", 17, "S", "Bonne"}, {"Omrani", "Fouad", 19, "S", "Très bonne"}, {"Berrada", "Amine", 20, "S", "Très bonne"}, {"Rafik", "Basma", 19, "L", "Moyenne"}, {"Sfendla", "Rime", 18, "ES", "Bonne"}, {"Semlali", "Mohammed", 19, "G", "Médiocre"}, {"Wahbi", "Salma", 17, "S", "Très bonne"}, {"Yacoubi", "Karim", 18, "L", "Très bonne"}}

Chapitre II – les tableaux statistiques (2)

↪ La matrice de données n'est pas lisible pour l'esprit humain

➔ ***Présentation des données dans un tableau***

Tab 1 : Résultat de l'enquête effectuée auprès des étudiants de sciences économiques

Noms	Prénoms	Age	Série du bac	Opinion sur l'architecture
Alaoui	Fatima	18	L	Très bonne
Otmani	Samira	17	S	Bonne
Omrani	Fouad	19	S	Très bonne
Berrada	Amine	20	S	Très bonne
Rafik	Basma	19	L	Moyenne
Sfendla	Rime	18	ES	Bonne
Semlali	Mohammed	19	G	Médiocre
Wahbi	Salma	17	S	Très bonne
Yacoubi	Karim	18	L	Très bonne

Chapitre II – les tableaux statistiques (3)

- **Exemple 2** : enquête auprès d'un échantillon de 56 familles marocaines sur le nombre d'enfant par ménage
- ➔ Données brutes : 0 0 0 1 1 1 1 1 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 4 4 4 4 5 5 5 5 5 5 5 5 5 6 6 6 6 6 6 7 7 8 9
- Les données brutes ne sont pas lisibles
- Regroupement des données dans un tableau pour faciliter le traitement et les interprétations

Chapitre II – les tableaux statistiques (4)

Nombre d'enfants par famille observé dans un échantillon de 56 familles

Modalités x_i
prises par
la variable x

<i>Nombre d'enfants/famille (x_i)</i>	<i>Effectif (n_i)</i>
0	3
1	5
2	8
3	7
4	14
5	9
6	6
7	2
8	1
9	1
Total	56

Nombre
d'individus
correspondant
à chaque
modalité

↓
Effectif ou
fréquence
absolue

Chapitre II – les tableaux statistiques (5)

- La première colonne du tableau reprend les différentes **modalités** (x_i) prises par la variable ou le caractère nombre d'enfants/ménage
 - La deuxième colonne présente les **effectifs** (n_i) (**fréquences absolues**) : le nombre d'individus correspondant à chaque modalité du caractère
 - Chaque case du tableau dénombre les individus considérés comme équivalents face au phénomène étudié
- ↪ L'ensemble des modalités et des effectifs d'un caractère forment une **distribution statistique** ou une **série statistique**

$$\textit{Distribution statistique} = \{x_i, n_i\}$$

Chapitre II – les tableaux statistiques (6)

- La présentation d'un tableau statistique doit respecter des principes généraux :
 - ➔ Le tableau doit porter un **titre** précisant son contenu : le phénomène étudié, la façon dont il est étudié, le lieu, la date, etc.
 - ➔ Le tableau doit porter des **intitulés** de lignes et de colonnes clairement définis
 - ➔ Le tableau doit préciser les **unités utilisés** : ne pas confondre le mètre avec le mètre carré, le millier avec le million, le DH avec l'Euro, etc
 - ➔ Le tableau doit préciser la **source** des informations lorsque les données sont empruntées à une publication ou à un organisme
- Les tableaux statistiques peuvent être à **une** ou à **plusieurs dimensions**
 - ➔ À « **une dimension** » si un seul caractère est étudié (nombre d'enfants/ménage)
 - ➔ À « **deux dimensions** » si l'on retient deux caractères (nombre et sexe des enfants/ménage)

Chapitre II – les tableaux statistiques (7)

Les tableaux à un caractère

- Considérons une population statistique de n individus décrite selon le caractère x dont les k modalités sont $x_1, x_2, \dots, x_i, \dots, x_k$

Modalités (x_i)	Effectif (n_i)
x_1	n_1
x_2	n_2
.	.
.	.
.	.
x_i	n_i
.	.
.	.
.	.
x_k	n_k
Total	n

- n_i représente le nombre d'individus, appelé « effectif partiel » ou « fréquence absolue », présentant la modalité x_i
- La somme des « effectifs partiels » n_i est « l'effectif total » n de la population

$$\sum_{i=1}^k n_i = n$$

Chapitre II – les tableaux statistiques (8)

Les tableaux à un caractère

- La « **fréquence relative** » ou « **fréquence** » f_i est la proportion d'individus présentant la même modalité dans la population

- La « **fréquence** » f_i est obtenue en divisant chaque effectif par l'effectif total

$$f_i = \frac{n_i}{n}$$

- La « **fréquence** » f_i peut être exprimée en pourcentage %

$$f_i \% = \frac{n_i}{n} * 100$$

- La somme des **fréquences relatives** f_i est égale à 1 et la somme des fréquences exprimées en % est égale à 100

→ **Démonstration**

Chapitre II – les tableaux statistiques (9)

Les tableaux à un caractère

↪ Le tableau statistique initiale se présentera sous la forme suivante :

Modalités (x_i)	Effectifs (n_i)	Fréquences (f_i)
x_1	n_1	f_1
x_2	n_2	f_2
x_3	n_3	f_3
.	.	.
.	.	.
.	.	.
x_i	n_i	f_i
.	.	.
.	.	.
.	.	.
x_k	n_k	f_k
Σ	n	1

Chapitre II – les tableaux statistiques (10)

Les tableaux à un caractère

Exemple d'application : Nombre d'enfants par famille

<i>Nombre d'enfants/famille (x_i)</i>	<i>Effectif (n_i)</i>
0	3
1	5
2	8
3	7
4	14
5	9
6	6
7	2
8	1
9	1
<i>Total</i>	<i>56</i>

1. Compléter le tableau en calculant les fréquences relatives et les fréquences en pourcentages
2. Interpréter les résultats

Chapitre II – les tableaux statistiques (11)

Les tableaux à caractères qualitatifs

- Les tableaux à caractères qualitatifs ne posent pas de problèmes particuliers
 1. Tableaux de caractère à modalités nominales
 2. Tableaux de caractère à modalités ordinales

1. Les tableaux de caractère à modalités nominales

→ Exemple1 : Répartition des salariés de l'entreprise M selon la CSP au 31/12/06

Modalités des CSP (x_j)	Effectifs (n_j)	Fréquences (f_j)	f_j en %
Cadres supérieurs	10		
Contremaîtres	5		
employés	30		
Ouvriers spécialisés	90		
Autres catégories	5		
Total	140		

Chapitre II – les tableaux statistiques (12)

Les tableaux à caractères qualitatifs

→ **Exemple 2** : Répartition des étudiants du groupe A selon leur lieu de naissance

Lieu de naissance (x_i)	Effectifs (n_i)	Fréquences (f_i)	f_i en %
Casablanca	98		
Mohammedia	53		
Rabat	47		
Kénitra	32		
Autres	20		
Total	250		

Chapitre II – les tableaux statistiques (13)

Les tableaux à caractères qualitatifs

2. Les tableaux de caractère à modalités ordinales

→ **Exemple** : enquête effectuée auprès d'un échantillon de 9 étudiants de sciences économiques sur leur opinion concernant l'architecture de la faculté

Opinions (x_i)	Effectifs (n_i)	Fréquences (f_i)	f_i en %
Très bonne	5		
Bonne	2		
Moyenne	1		
Médiocre	1		
Total	9		

Chapitre II – les tableaux statistiques (14)

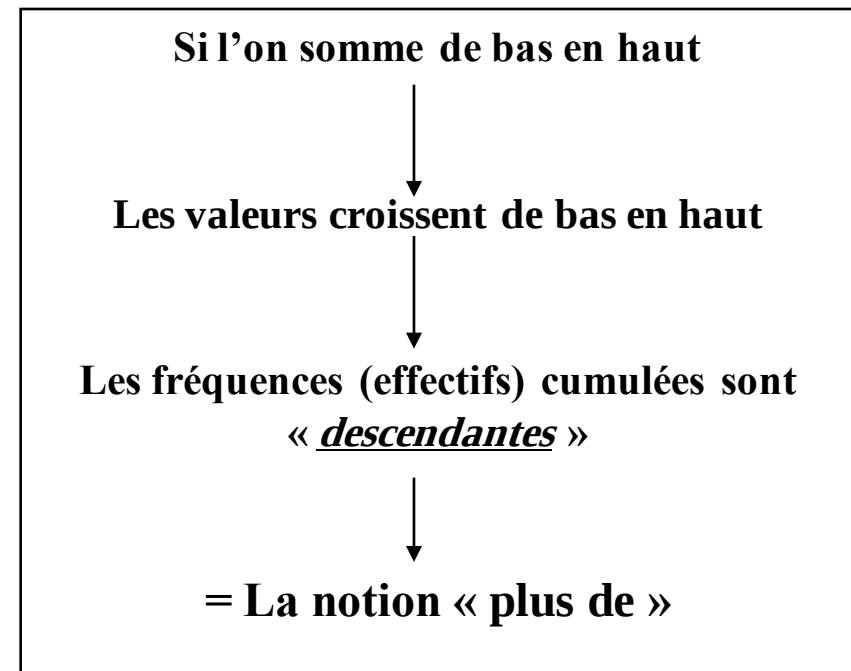
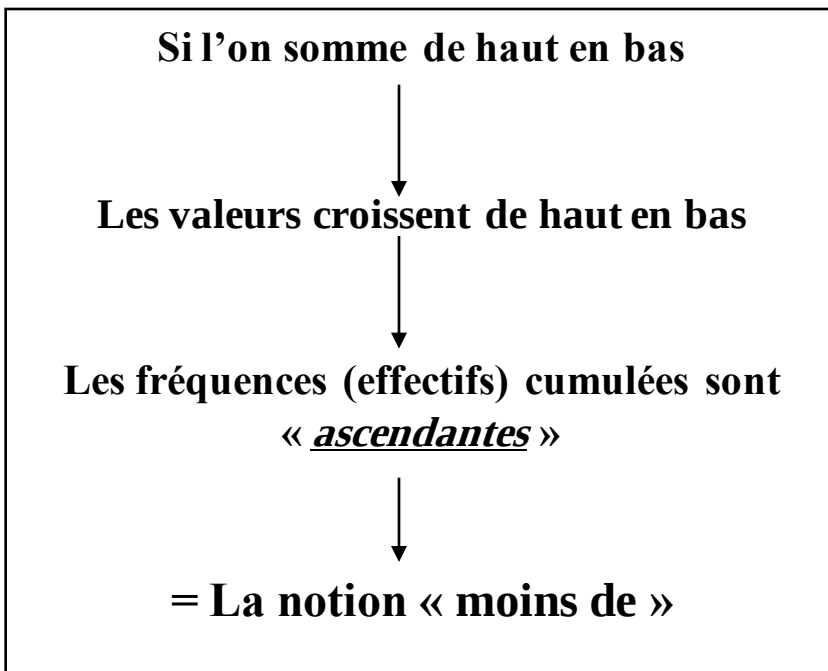
Les tableaux à caractères quantitatifs

- Les tableaux à caractères quantitatifs peuvent contenir plus d'informations que les tableaux à caractères qualitatifs :
 - Effectifs cumulés
 - Fréquences cumulées
- **Les effectifs cumulés** notés $N(x)$
 - Exemple nombre d'enfants par famille : **Combien de familles ont plus de quatre enfants? Combien de familles ont moins de quatre enfants?**
- **Les fréquences cumulées** notées $F(x)$
 - **Quelle est la proportion de familles ayant plus de quatre enfants? Quelle est la proportion de familles ayant moins de quatre enfants?**

Chapitre II – les tableaux statistiques (15)

Les tableaux à caractères quantitatifs

- Le **calcul** des effectifs cumulés et des fréquences cumulées se fait en **cumulant** (sommant) les **effectifs** et les **fréquences relatives** dans une colonne du tableau



Chapitre II – les tableaux statistiques (16)

Les tableaux à caractères quantitatifs

1. Cas de caractères quantitatifs discrets

→ Exemple : Nombre d'enfants (x_i) observés dans un échantillon de 55 familles

<i>Nombre d'enfants</i>	<i>Ménages</i>	f_i	<i>Eff. cumulés « moins de »</i>	<i>Eff. cumulés « plus de »</i>	<i>Fréq. cumulées « moins de »</i>	<i>Fréq. cumulées « plus de »</i>
0	3					
1	4					
2	8					
3	7					
4	14					
5	9					
6	6					
7	2					
8	1					
9	1					
Totaux	55					

Chapitre II – les tableaux statistiques (17)

Les tableaux à caractères quantitatifs

Nombre d'enfants (x_i) observés dans un échantillon de 55 familles

<i>Nombre d'enfants</i>	<i>Ménages</i>	f_i	<i>Eff. cumulés « moins de »</i>	<i>Eff. cumulés « plus de »</i>	<i>Fréq. cumulées « moins de »</i>	<i>Fréq. cumulées « plus de »</i>
0	3	0,05	3	55	0,05	1
1	4	0,07	3+4= 7	52	0,12	0,95
2	8	0,15	7+8= 15	48	0,27	0,88
3	7	0,13	15+7= 22	40	0,4	0,73
4	14	0,25	36	33	0,65	0,60
5	9	0,16	45	19	0,81	0,35
6	6	0,11	51	4+6= 10	0,92	0,19
7	2	0,04	53	2+2= 4	0,96	0,08
8	1	0,02	54	1+1= 2	0,98	0,04
9	1	0,02	55	1	1	0,02
Totaux	55	1	-	-	-	-

Chapitre II – les tableaux statistiques (18)

Les tableaux à caractères quantitatifs

- Il y a 15 ménages dans l'échantillon qui ont « moins de » 3 enfants. On peut dire aussi qu'il y a 15 ménages dans l'échantillon qui ont « au plus » 2 enfants
- Il y a 40 ménages dans l'échantillon qui ont « plus de » 2 enfants. On peut dire aussi qu'il y a 40 ménages dans l'échantillon qui ont « au moins » 3 enfants
- 40% des ménages de l'échantillon ont « moins de » 4 enfants ou « au plus » 3 enfants
- 60% des ménages de l'échantillon ont « plus de » 3 enfants ou « au moins » 4 enfants

Chapitre II – les tableaux statistiques (19)

Les tableaux à caractères quantitatifs

2. Cas de caractères quantitatifs continus

→ Exemple : Répartition des salaires mensuels d'une entreprise X au 31/12/06

<i>Classes de Salaires</i>	n_i	f_i	<i>Eff. cumulés « moins de »</i>	<i>Eff. cumulés « plus de »</i>	<i>Fréq. cumulées « moins de »</i>	<i>Fréq. cumulées « plus de »</i>
[6000-7000[	10					
[7000-9000[	50					
[9000-10 000[	200					
[10 000-13 000[	20					
[13 000-17 000[	10					
17 000 et +	5					
Totaux	295	1	-	-	-	-

Chapitre II – les tableaux statistiques (20)

Les tableaux à caractères quantitatifs

■ Interprétation des résultats

- 88% des salariés gagnent moins de 10000 DH par mois (260 personnes)
- 80% des salariés de gagnent plus de 9000 DH par mois (235 personnes)

■ Remarque importante

- Le calcul des fréquences et des effectifs cumulés n'est pas affecté par l'amplitudes des classes

Chapitre II – les tableaux statistiques (21)

Les tableaux à deux caractères

- Une population statistique peut être décrite à l'aide de deux caractères simultanément
 - ➔ **Ex 1** : la population des ménages peut être décrite selon son **revenu** et ses **dépenses** simultanément
 - ➔ **Ex 2** : la population active marocaine peut être décrite en fonction de la **CSP** et du **niveau de formation**



Les tableaux statistiques correspondant sont à deux dimensions

- ↪ Les tableaux de contingence ou croisés dynamiques ou à double entrées

Chapitre II – les tableaux statistiques (22)

Présentation générale des tableaux de contingence

■ Considérons une population statistique décrite selon deux caractères :

→ Un caractère X dont les n modalités x_i sont $x_1, x_2, \dots, x_j, \dots, x_n$

→ Un caractère Y dont les k modalités y_j sont $y_1, y_2, \dots, y_j, \dots, y_k$

Les k modalités de Y

y_j x_i	y_1	y_2	$\dots$	y_j	$\dots$	y_k	$n_{i.}$
x_1	n_{11}	n_{12}	$\dots$	n_{1j}	$\dots$	n_{1k}	$n_{1.}$
x_2	n_{21}	n_{22}	$\dots$	n_{2j}	$\dots$	n_{2k}	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_i	n_{i1}	n_{i2}	$\dots$	n_{ij}	$\dots$	n_{ik}	$n_{i.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_n	n_{n1}	n_{n2}	$\dots$	n_{nj}	$\dots$	n_{nk}	$n_{n.}$
$n_{.j}$	$n_{.1}$	$n_{.2}$	$\dots$	$n_{.j}$	$\dots$	$n_{.k}$	$n_{..}$

→ Les effectifs partiels apparaissent à l'intérieur du tableau

⇒ n_{ij} : effectif de la population présentant à la fois la modalité x_i et la modalité y_j

⇒ n_{ij} : l'indice de X « i » d'abord et de Y « j » ensuite

→ Les marges ou effectifs marginaux

⇒ $n_{i.}$: somme des effectifs de la $i^{\text{ème}}$ ligne, l'indice j variant de 1 à K est remplacé par « . »

⇒ $n_{.j}$: somme des effectifs de la modalité y_j , l'indice $i = 1$ à n est remplacé par « . »

Les n modalités de X

Chapitre II – les tableaux statistiques (23)

Présentation générale des tableaux de contingence

■ Le tableau de contingence obéit à une notation conventionnelle

1. Le tableau contient :

→ Dans la 1^{ère} colonne les n modalités $x_1, x_2, \dots, x_j, \dots, x_n$ du caractère X

→ Dans la 1^{ère} ligne les k modalités $y_1, y_2, \dots, y_j, \dots, y_k$ du caractère Y

2. L'effectif n_{ij} correspond à l'intersection d'une ligne i et d'une colonne j

→ L'effectif de la population présentant à la fois la modalité x_i et la modalité y_j

3. Pour les effectifs marginaux $n_{i.}$ et $n_{.j}$, on remplace l'indice qui varie par « . »

→ $n_{i.}$: somme des effectifs de la $i^{\text{ème}}$ ligne, $j=1, \dots, K$ est remplacé par « . »

→ $n_{.j}$: somme des effectifs de la $j^{\text{ème}}$ colonne, $i=1, \dots, n$ est remplacé par « . »

4. L'effectif général marginal de X est noté « $n_{i.}$ » et celui de Y « $n_{.j}$ »

5. L'effectif total du tableau est noté « $n_{..}$ » : il s'agit de l'effectif total de la population étudiée

Chapitre II – les tableaux statistiques (24)

Présentation générale des tableaux de contingence

- **Exemple** : répartition des salariés d'une entreprise X selon le sexe (x_i) et le niveau de formation (y_j)

$x_i \backslash y_j$	Féminin	Masculin	<i>Total</i> ($n_{i.}$)
Bac + 3	45	49	94
Bac + 5	16	11	27
Bac + 8	4	6	10
<i>Total</i> ($n_{.j}$)	65	66	$n_{..} = 131$

→ La marge $n_{.j}$ et la ligne du haut y_j donne la distribution marginale des salariés de l'entreprise selon leur « sexe »



**Distribution marginale
du caractère Y**

→ La marge $n_{i.}$ et la première colonne x_i donne la distribution marginale des salariés de l'entreprise selon leur « niveau de formation »

→ Distribution marginale du caractère X

Chapitre II – les tableaux statistiques (25)

Propriétés des tableaux de contingence

- a) Les modalités de x_i et y_j étant incompatibles et exhaustives, on peut écrire plusieurs séries d'égalités

→ Pour y_j

$$\sum_{j=1}^k n_{1j} = n_{1.} \longrightarrow n_{1.} \text{ représente le nombre d'individus présentant la modalité } x_1 \text{ de } X \text{ quelle que soit la modalité de } y$$

De façon générale :

$$\sum_{j=1}^k n_{ij} = n_{i.}$$

→ Pour x_i

$$\sum_{i=1}^n n_{i1} = n_{.1} \longrightarrow n_{.1} \text{ représente le nombre d'individus présentant la modalité } y_1 \text{ de } Y \text{ quelle que soit la modalité de } x$$

De façon générale :

$$\sum_{i=1}^n n_{ij} = n_{.j}$$

Chapitre II – les tableaux statistiques (26)

Propriétés des tableaux de contingence

- L'effectif total de la population $n_{..}$

- Apparaît à l'intersection de la dernière ligne et de la dernière colonne
- Est égal à la somme de la dernière ligne ou de la dernière colonne

$$n_{..} = \sum_{i=1}^n n_{i.} = \sum_{j=1}^k n_{.j}$$

En remplaçant $n_{i.}$ et $n_{.j}$ par les expressions précédentes, on obtient

$$n_{..} = \sum_{i=1}^n \sum_{j=1}^k n_{ij} = \sum_{j=1}^k \sum_{i=1}^n n_{ij}$$

Chapitre II – les tableaux statistiques (27)

Propriétés des tableaux de contingence

b) Les fréquences partielles

→ Rapport de l'effectif partiel sur l'effectif total

↪ La fréquence partielle des modalités x_i , y_j est égale à :

$$f_{ij} = \frac{n_{ij}}{n_{..}}$$

↪ *Proportion d'individus satisfaisant à la fois la modalité x_i et la modalité y_j*

→ La somme des fréquences partielles est égale à 1

⇒ *Démonstration*

Chapitre II – les tableaux statistiques (28)

Propriétés des tableaux de contingence

- **Exemple** : répartition des salariés d'une entreprise M selon le sexe (x_i) et le niveau de formation (y_j)

$x_i \backslash y_j$	Féminin	Masculin	<i>Total</i> ($n_{i.}$)
Bac + 3	45	49	94
Bac + 5	16	11	27
Bac + 8	4	6	10
<i>Total</i> ($n_{.j}$)	65	66	$n_{..} = 131$

1. Calculer f_{22} , f_{31} , f_{12}

2. Interpréter les résultats

$$f_{22} = \frac{11}{131}$$

8% des salariés sont des femmes de niveau Bac + 5

Chapitre II – les tableaux statistiques (29)

Les différentes distributions statistiques

- Plusieurs distributions statistiques peuvent être définies dans un tableau à double entrées
 - ➔ Les distributions marginales
 - ➔ Les distributions conditionnelles

1. Les distributions marginales

- ➔ Un tableau de contingence compte deux distributions marginales : la distribution marginale du caractère X et la distribution marginale du caractère Y

⇒ La distribution marginale du caractère X

- Est composée des modalités du caractère X et des effectifs correspondant quelles que soit les modalités du caractère Y

Chapitre II – les tableaux statistiques (30)

Les différentes distributions statistiques

⇒ La distribution marginale du caractère X est donnée par le tableau suivant

<i>Caractère</i>	<i>Effectifs marginaux</i>	<i>Fréquences marginales</i>
x_1	$n_{1.}$	$f_{1.}$
x_2	$n_{2.}$	$f_{2.}$
$\vdots$	$\vdots$	$\vdots$
x_i	$n_{i.}$	$f_{i.}$
$\vdots$	$\vdots$	$\vdots$
x_n	$n_{n.}$	$f_{n.}$
Total	$n=n_{..}$	1

→ On peut calculer les « **fréquences marginales** » : rapport de l'effectif marginal sur l'effectif total

$$f_{i.} = \frac{n_{i.}}{n_{..}}$$

Démonstration

Chapitre II – les tableaux statistiques (31)

Les différentes distributions statistiques

⇒ La distribution marginale du caractère Y

- Est composée des modalités du caractère Y et des effectifs correspondant quelles que soit les modalités du caractère X

- **La fréquence marginale** de la modalité y_j est égale à : $f_{.j} = \frac{n_{.j}}{n_{..}}$

<i>Caractère</i>	<i>Effectifs marginaux</i>	<i>Fréquences marginales</i>
y_1	$n_{.1}$	$f_{.1}$
y_2	$n_{.2}$	$f_{.2}$
$\vdots$	$\vdots$	$\vdots$
y_i	$n_{.j}$	$f_{.j}$
$\vdots$	$\vdots$	$\vdots$
y_k	$n_{.k}$	$f_{.k}$
Total	$n=n_{..}$	1

Démonstration

Chapitre II – les tableaux statistiques (32)

Les différentes distributions statistiques

- **Exemple d'application** : répartition des salariés d'une entreprise M selon le sexe (x_i) et le niveau de formation (y_j)

$x_i \backslash y_j$	Féminin	Masculin	<i>Total</i> ($n_{i.}$)
Bac + 3	45	49	94
Bac + 5	16	11	27
Bac + 8	4	6	10
<i>Total</i> ($n_{.j}$)	65	66	$n_{..} = 131$

Distribution marginale des effectifs des salariés en fonction de leur niveau de formation

Distribution marginale des effectifs des salariés en fonction de leur sexe

Chapitre II – les tableaux statistiques (32)

Les différentes distributions statistiques

2. Les distributions conditionnelles

→ Deux séries de distributions conditionnelles

⇒ Celle du caractère X conditionnellement au caractère Y

⇒ Celle du caractère Y conditionnellement au caractère X

⇒ Distributions conditionnelles du caractère X liées par $y_j, j=1, \dots, k$

➤ Ce sont les modalités de X et des effectifs de chacune de ces modalités dans la sous population présentant la modalité y_j de Y

➤ **Exemple** : répartition de la sous population des **femmes** de l'entreprise M selon leur niveau de formation

$x_j \backslash y_j$	Féminin
Bac + 3	45
Bac + 5	16
Bac + 8	4
Total ($n_{.j}$)	65

Chapitre II – les tableaux statistiques (33)

Les différentes distributions statistiques

⇒ Distribution conditionnelle du caractère X liée par y_j ($j=1$ à k) est la suivante :

<i>Caractère</i>	<i>Effectifs de y_j</i>	<i>Fréquences conditionnelles</i>
x_1	n_{1j}	$f_{1/j}$
x_2	n_{2j}	$f_{2/j}$
$\vdots$	$\vdots$	$\vdots$
x_i	n_{ij}	$f_{i/j}$
$\vdots$	$\vdots$	$\vdots$
x_n	n_{nj}	$f_{n/j}$
Total	$n_{.j}$	1

→ On peut calculer la « **fréquence conditionnelle** » de la modalité x_i de X sous condition que $Y=y_j$: **proportion d'individus présentant la modalité x_i parmi les individus qui présentent uniquement la modalité y_j**

$$f_{i/j} = f_i^j = \frac{n_{ij}}{n_{.j}}$$

Démonstration

Chapitre II – les tableaux statistiques (34)

Les différentes distributions statistiques

⇒ Distributions conditionnelles du caractère Y liées par $x_i, i=1, \dots, n$

- Ce sont les modalités de Y et des effectifs de chacune de ces modalités dans la sous population présentant la modalité x_i de X
- **Exemple** : répartition de la sous population de l'entreprise M ayant un niveau de formation « Bac+3 » selon le sexe

y_j x_i	Féminin	Masculin	Total ($n_{i.}$)
Bac + 3	45	49	94

- La « **fréquence conditionnelle** » de la modalité y_j de y sous condition que $x = x_i$: **proportion d'individus présentant la modalité y_j parmi les individus qui présentent uniquement la modalité x_i**

$$f_{j/i} = f_j^i = \frac{n_{ij}}{n_{i.}}$$

Chapitre II – les tableaux statistiques (35)

Les différentes distributions statistiques

⇒ Distribution conditionnelle du caractère Y liées par x_i ($i=1, \dots, n$) est la suivante

<i>Caractère</i>	<i>Effectifs de x_i</i>	<i>Fréquences conditionnelles</i>
y_1	n_{i1}	$f_{1/i}$
y_2	n_{i2}	$f_{2/i}$
$\vdots$	$\vdots$	$\vdots$
y_j	n_{ij}	$f_{j/i}$
$\vdots$	$\vdots$	$\vdots$
y_k	n_{ik}	$f_{k/i}$
Total	$n_{i.}$	1

Démonstration

Chapitre II – les tableaux statistiques (36)

Les différentes distributions statistiques

Exemple d'application : répartition des salariés d'une entreprise M selon le sexe (x_i) et le niveau de formation (y_j)

$x_i \backslash y_j$	Féminin	Masculin	Total ($n_{i.}$)
Bac + 3	45	49	94
Bac + 5	16	11	27
Bac + 8	4	6	10
Total ($n_{.j}$)	65	66	$n_{..} = 131$

1. Calculer f_{ij} : $f_{2/2}$, $f_{3/1}$, $f_{1/2}$

2. Calculer f_{ji} : $f_{2/2}$, $f_{3/1}$, $f_{1/2}$

3. Interpréter les résultats

$$1. f_{2/2} = \frac{n_{22}}{n_{.2}} = \frac{11}{66} = 0,17 \rightarrow$$

Parmi les 66 salariés « **hommes** » de l'entreprise, 11 ont un niveau de formation « **Bac+5** »

17% des salariés « **hommes** » de l'entreprise ont un niveau de formation « **Bac+5** »

Chapitre II – les tableaux statistiques (37)

Les différentes distributions statistiques

3. Relation entre les fréquences marginales et les fréquences conditionnelles

→ On peut démontrer que le produit des **fréquences marginales** par les **fréquences conditionnelles** est égal aux **fréquences partielles**

$$f_{i.} \times f_{j/i} = f_{ij}$$

et

$$f_{.j} \times f_{i/j} = f_{ij}$$

→ *Démonstrations*

Chapitre II – les tableaux statistiques (37)

Conclusion

- Le chapitre II s'est intéressé à la présentation de données statistiques brutes de manière synthétique dans des tableaux statistiques
 - Présentation « lisible » et « compréhensible » des données
- Les données peuvent également être représentées sous forme « graphique »
 - Les représentations graphiques reproduisent sous une autre forme les données fournies par les tableaux statistiques
 - Courbes, diagrammes, histogrammes, dessins, etc.

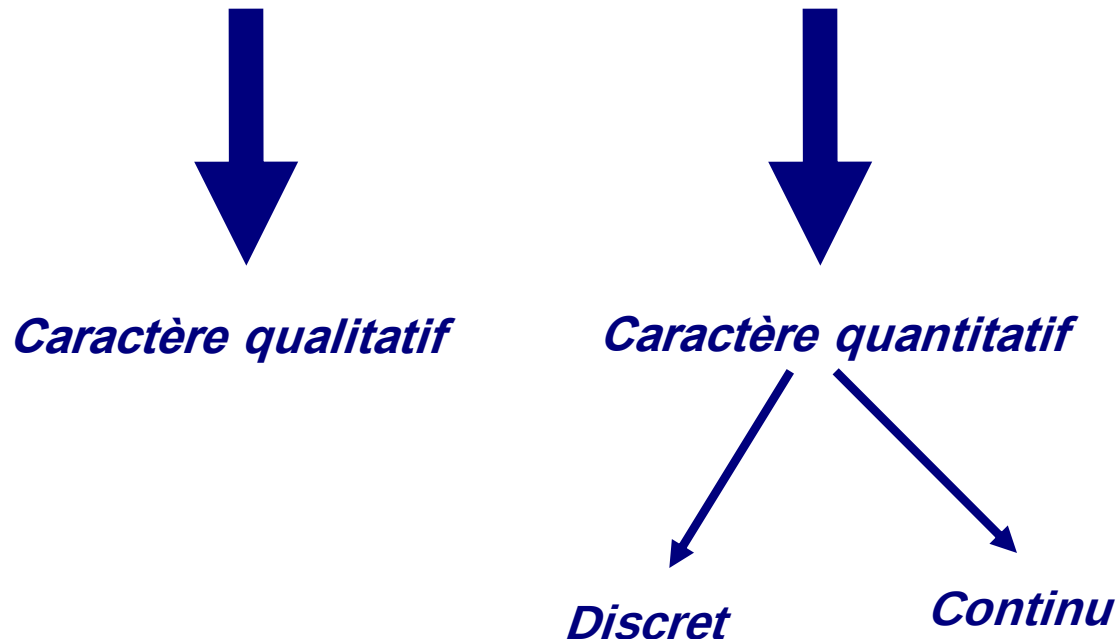
Chapitre III – Les représentations graphiques (1)

- Les graphiques permettent de donner une synthèse visuelle de la distribution d'une variable
- Les graphes apparaissent comme plus « parlants » que les tableaux
 - Ils donnent, au sens propre, *une image* des réalités observées
- Les représentations graphiques sont spécifiques à un type de variables ou de caractères
 - Qualitatif : ordinal / nominal
 - Quantitatif : discret / continu
- ✚ **Section 1** : Les représentations graphiques des distributions statistiques à une dimension
- ✚ **Section 2** : Les représentations graphiques des distributions statistiques à deux dimensions

Chapitre III – Les représentations graphiques (2)

Représentations des distributions à une dimension

- Le choix des représentations graphiques dépend de la nature du caractère statistique étudié



Chapitre III – Les représentations graphiques (3)

Représentations des caractères qualitatifs

- Les variables qualitatives peuvent être représentées graphiquement de différentes manières
 - ➔ *Diagrammes en bâtons*
 - ➔ *Diagrammes en barres (ou en tuyaux d'orgue)*
 - ➔ *Diagrammes circulaires (ou en camembert ou en secteurs)*

1. Diagrammes en bâtons

- Un diagramme en bâtons est constitué d'une suite de « bâtons »
- A chaque modalité x_i du caractère, on associe un « bâton » de longueur h_i
- La longueur h_i doit être proportionnelle à la fréquence f_i ou à l'effectif n_i
- Les bâtons peuvent être verticaux ou horizontaux

Chapitre III – Les représentations graphiques (4)

Représentations des caractères qualitatifs

- Exemple : Répartition des salariés de l'entreprise X selon la CSP

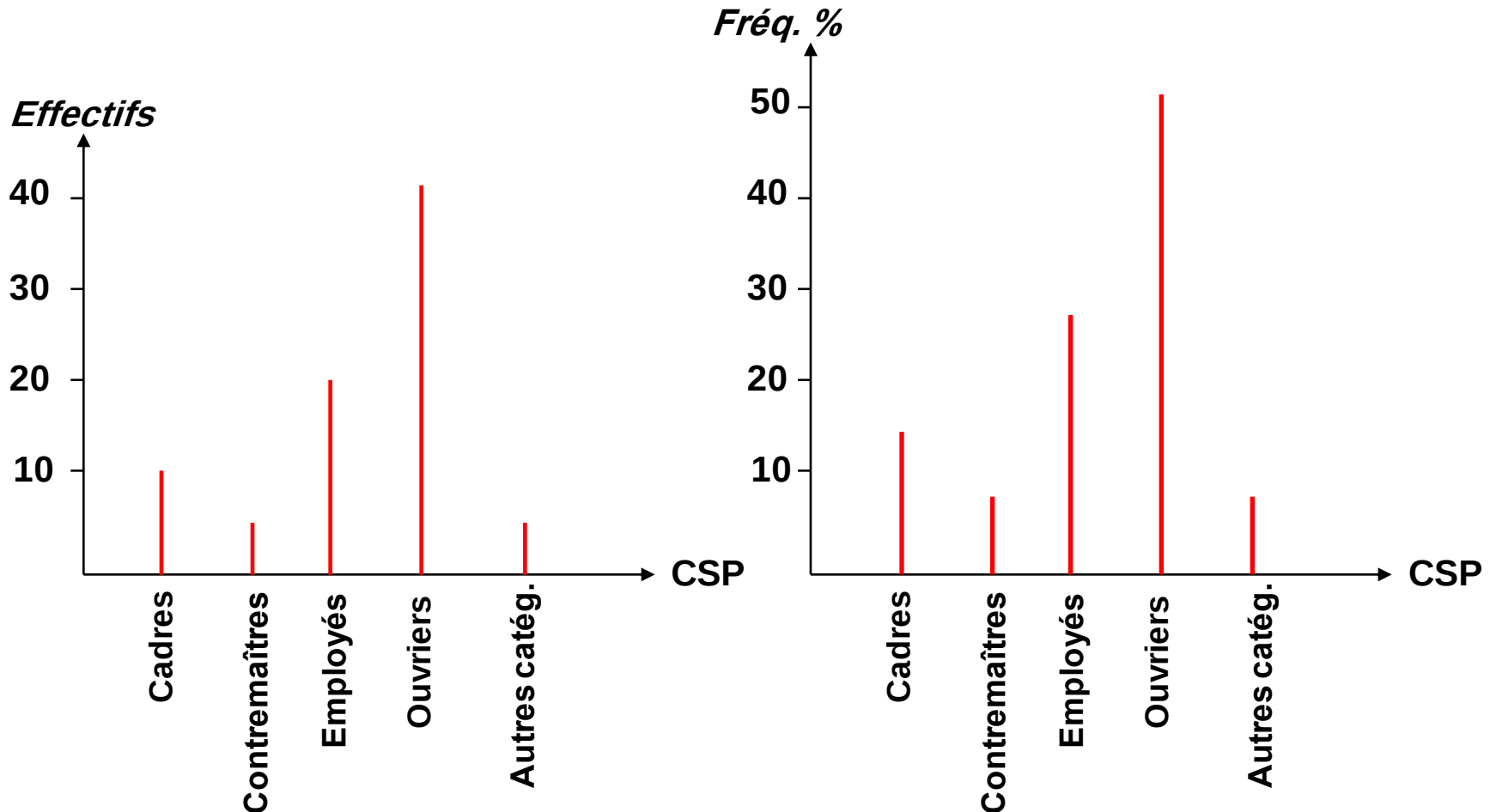
CSP (x_i)	Effectifs (n_i)	Fréquences f_i %
Cadres supérieurs	10	12,5%
Contremaîtres	5	6,25%
Employés	20	25%
Ouvriers spécialisés	40	50%
Autres catégories	5	6,25%
Total	80	100%

➔ Représenter graphiquement la distribution étudiée

Chapitre III – Les représentations graphiques (5)

Représentations des caractères qualitatifs

Répartition des salariés de l'entreprise selon la CSP

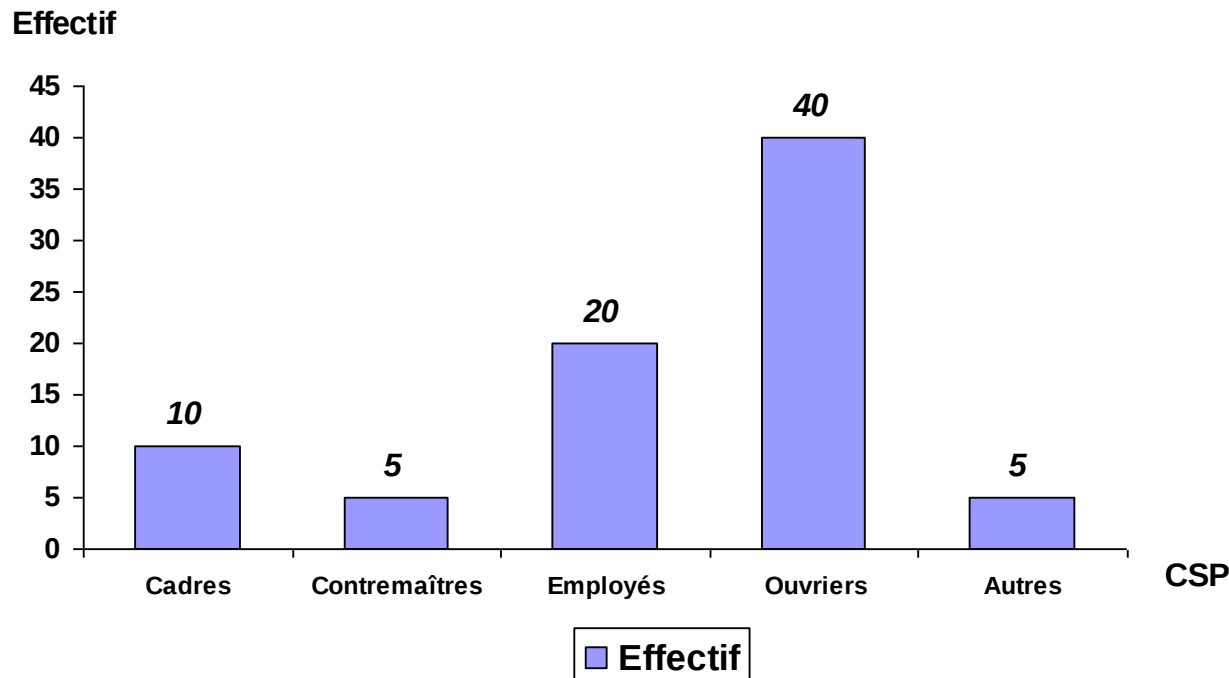


Chapitre III – Les représentations graphiques (6)

Représentations des caractères qualitatifs

2. Diagrammes en barres

→ Même principe que pour les diagrammes en bâtons



Chapitre III – Les représentations graphiques (7)

Représentations des caractères qualitatifs

3. Diagrammes circulaires (camembert ou secteurs)

- Cercle divisé en secteurs représentant l'ensemble de la population
 - Les différentes modalités du caractère sont représentées par des secteurs dont la surface est proportionnelle aux effectifs ou fréquences
- L'angle de chaque secteur α_i est proportionnel à la fréquence f_i : $\alpha_i = 360 \times f_i$

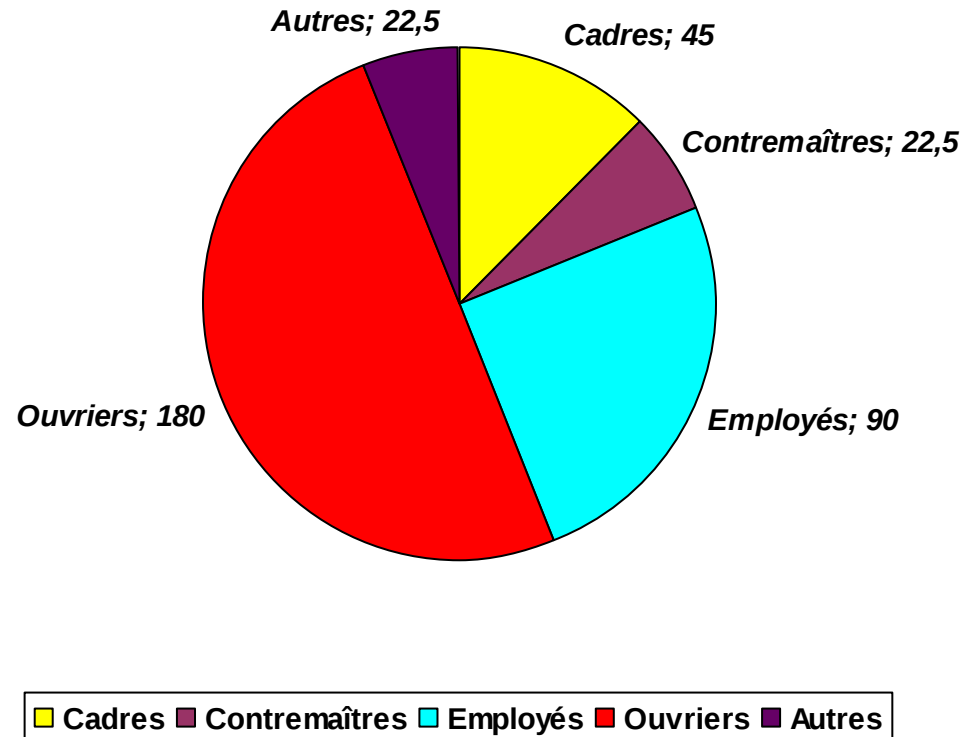
- **Exemple**: Répartition des salariés de l'entreprise X selon la CSP

CSP (x_i)	n_i	f_i	Angle α_i (en d°)
Cadres	10	0,125	45
Contremaîtres	5	0,0625	22,5
Employés	20	0,25	90
Ouvriers	40	0,50	180
Autres	5	0,0625	22,5

Chapitre III – Les représentations graphiques (8)

Représentations des caractères qualitatifs

Répartition par secteurs des salariés de l'entreprise selon la CSP



Chapitre III – Les représentations graphiques (9)

Représentations des caractères quantitatifs

1. Représentation graphique des caractères quantitatifs discrets

- Représentation d'une distribution des fréquences (ou effectifs)
- Représentation d'une distribution des fréquences cumulées (ou effectifs cumulés)

a) Représentation d'une distribution de fréquences (ou effectifs)

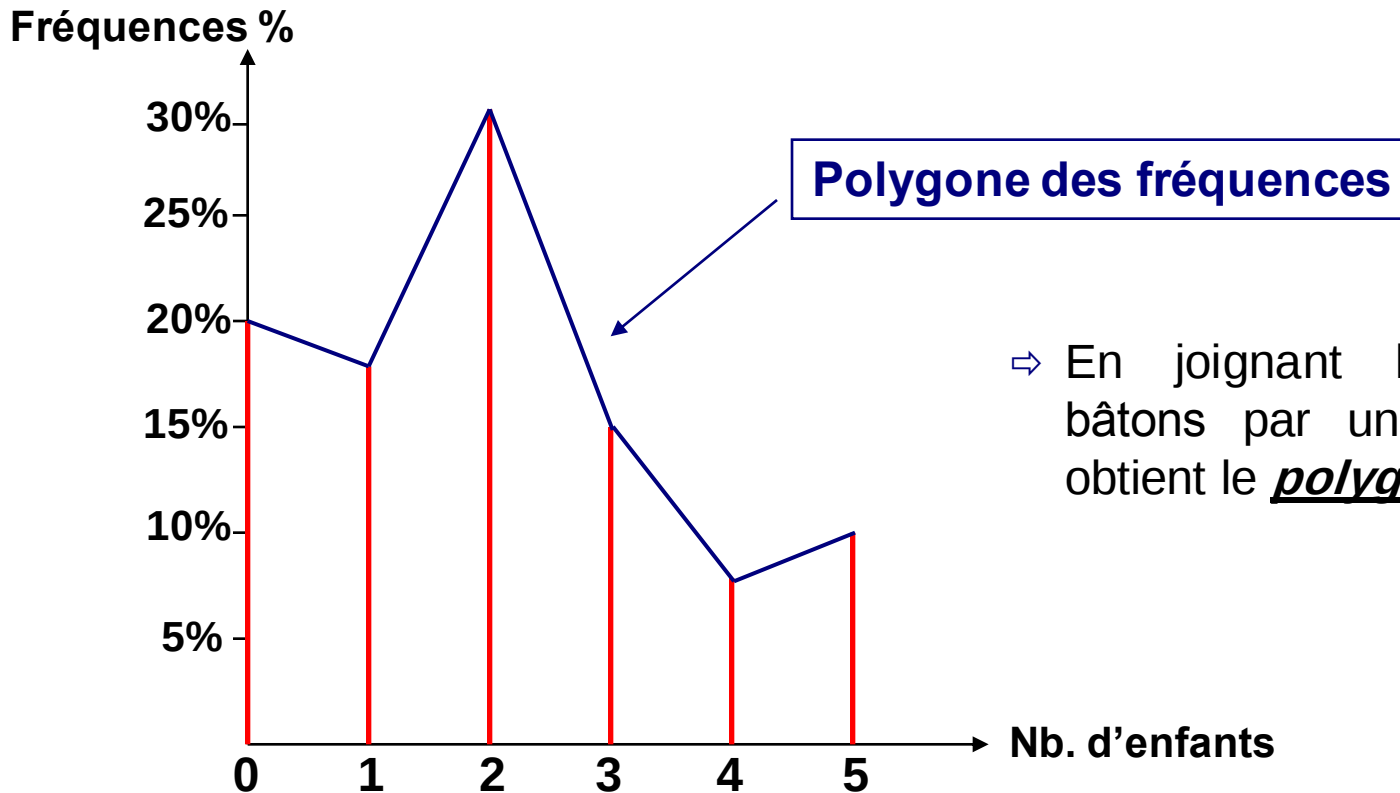
- Les fréquences (ou effectifs) sont représentées par les diagrammes en bâtons
 - Exemple: Nombre d'enfants des 40 salariés d'une entreprise

Nb. d'enfants	Effectifs	Fréquences %
0	8	20%
1	7	17,5%
2	12	30%
3	6	15%
4	3	7,5%
5	4	10%
Total	40	100%

Chapitre III – Les représentations graphiques (10)

Représentations des caractères quantitatifs

Distribution des fréquences des salariés selon leur nombre d'enfants



⇒ En joignant les sommets des bâtons par une ligne brisée, on obtient le **polygone de fréquences**

Chapitre III – Les représentations graphiques (11)

Représentations des caractères quantitatifs

b) Représentation d'une distribution de fréquences cumulées (ou eff. cum.)

- Pour représenter une distribution de fréquences (effectifs) cumulées, il faut d'abord définir la ***fonction de répartition*** $F(x)$
- Considérons une population statistique décrite selon un caractère quantitatif discret X dont les n modalités x_i sont : $x_1, x_2, \dots, x_i, \dots, x_n$
 - ⇒ où $x_1 < x_2 < x_i < \dots < x_n$
- La ***fonction de répartition*** $F(x)$ est définie comme suit :

$$F(x) = \begin{cases} 0, & \text{si } x < x_1 \\ f_i, & \text{si } x_i \leq x < x_{i+1}, (i=1, \dots, n) \\ 1, & \text{si } x \geq x_n \end{cases}$$

Chapitre III – Les représentations graphiques (12)

Représentations des caractères quantitatifs

- Il y a plusieurs définitions possibles d'une **fonction de répartition** $F(x)$ selon que sont considérées les fréquences (effectifs) cumulées « **moins de** », « **plus de** »

→ $F(x)$ représentant les fréquences cumulées « moins de » x :

$$F(x) = \sum_{x_i < x} f_i \longrightarrow \text{La somme des fréquences correspondant à toutes les modalités du caractère } X \text{ inférieures à } x$$

→ $F(x)$ représentant les fréquences cumulées « plus de » x :

$$F(x) = \sum_{x_i > x} f_i \longrightarrow \text{La somme des fréquences correspondant à toutes les modalités du caractère } X \text{ supérieures à } x$$

Chapitre III – Les représentations graphiques (13)

Représentations des caractères quantitatifs

- La *fonction de répartition* $F(x)$ (ou $N(x)$) est représentée par la courbe cumulative des fréquences (effectifs)

→ Exemple : Nombre d'enfants des 40 salariés d'une entreprise

Nb. d'enfants	Effectifs	Fréq. %	$F(x)$ « moins de »	$F(x)$ « plus de »
0	8	20%	20%	100%
1	7	17,5%	37,5%	80%
2	12	30%	60,5%	62,5%
3	6	15%	82,5%	32,5%
4	3	7,5%	90%	17,5%
5	4	10%	100%	10%
<i>Total</i>	<i>40</i>	<i>100%</i>	-	-

Chapitre III – Les représentations graphiques (14)

Représentations des caractères quantitatifs

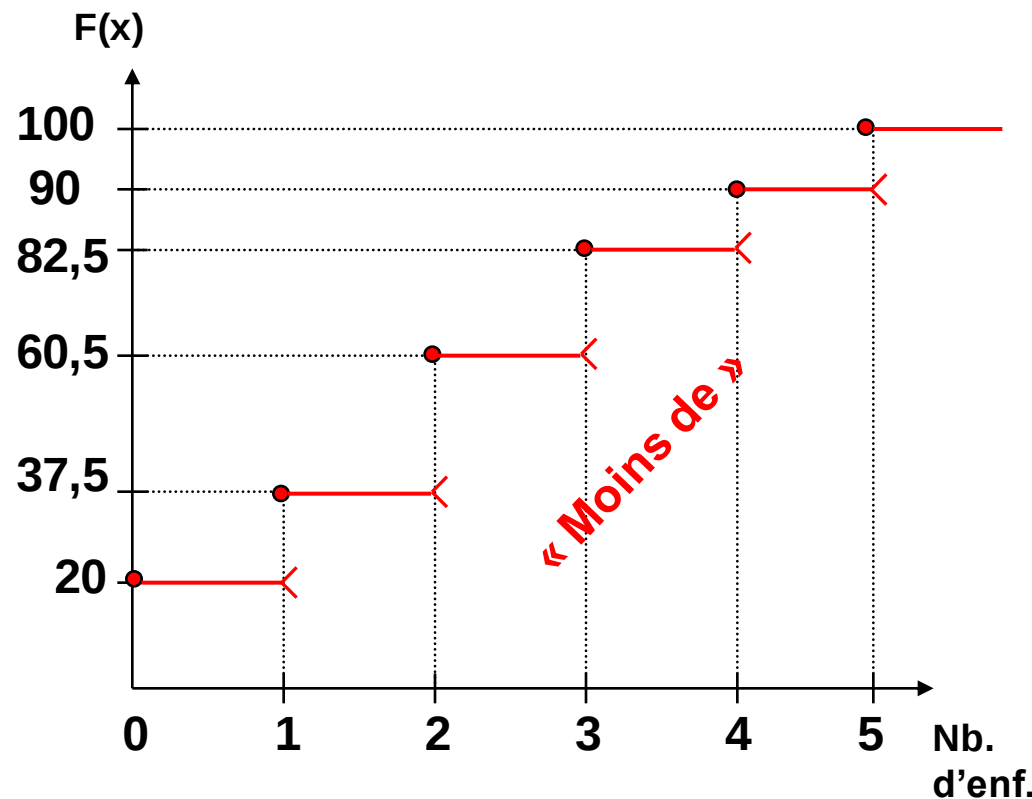
Représentation graphique de la courbe cumulative croissante

- La fonction $F(x)$ étant définie sur $\{0,1,2,3,4,5\}$, on commence par représenter les points dont on connaît les coordonnées
- Entre chaque modalité du caractère, $F(x)$ est constante

$$F(x) = 20\% \quad \forall x < 1$$

$$F(x) = 37,5\% \quad \forall 1 \leq x < 2$$

$$F(x) = 60,5\% \quad \forall 2 \leq x < 3$$



Chapitre III – Les représentations graphiques (15)

Représentations des caractères quantitatifs

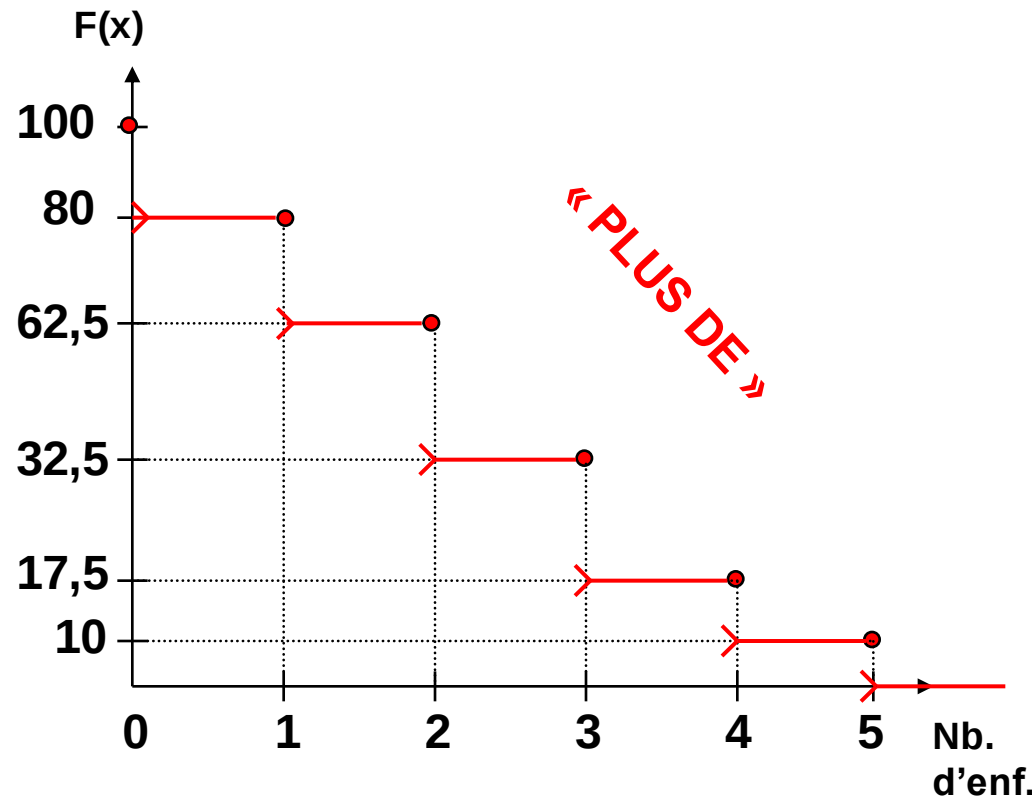
Représentation graphique de la courbe cumulative décroissante

- La fonction $F(x)$ étant définie sur $\{0,1,2,3,4,5\}$, on commence par représenter les points dont on connaît les coordonnées
- Entre chaque modalité du caractère, $F(x)$ est constante

$$F(x) = 80\% \quad \forall 0 < x \leq 1$$

$$F(x) = 62,5\% \quad \forall 1 < x \leq 2$$

$$F(x) = 32,5\% \quad \forall 2 < x \leq 3$$



Chapitre III – Les représentations graphiques (16)

Représentations des caractères quantitatifs

2. Représentation graphique des caractères quantitatifs continus

- Représentation d'une distribution des fréquences (ou effectifs)
- Représentation d'une distribution des fréquences cumulées (ou effectifs cumulés)

a) Représentation d'une distribution de fréquences (ou effectifs)

- Les fréquences (ou effectifs) des variables quantitatives continues sont représentées graphiquement par **les histogrammes**
 - À chaque classe de valeurs, on fait correspondre ***un rectangle*** dont l'aire est proportionnelle à la fréquence (ou l'effectif) de chaque classe
 - Deux cas de figures doivent être envisagés selon que les amplitudes de classes sont égales ou inégales
 - ⇒ **Cas de classes d'amplitudes égales**
 - ⇒ **Cas de classes d'amplitudes inégales**

Chapitre III – Les représentations graphiques (17)

Représentations des caractères quantitatifs

⇒ Cas de classes d'amplitudes égales

- Sur l'axe des abscisses, sont portées les limites des classes
- Sur l'axe des ordonnées, sont portées les fréquences (ou effectifs) correspondant à chaque classe

→ *Chaque fréquence (ou effectif) est représentée par un rectangle dont la base représente l'amplitude de classe et dont la hauteur est proportionnelle à la fréquence (ou effectif)*

→ Si :

$\underline{a}_i$: représente l'amplitude constante des classes

$\underline{f}_i$: représente la fréquence correspondant à chaque classe

↪ La surface du rectangle correspondant à la classe i est : S_i

$$S_i = a_i \times f_i$$

Chapitre III – Les représentations graphiques (18)

Représentations des caractères quantitatifs

- **Exemple** : Salaires des 50 employés de l'entreprise « X » en DH au 31/11/2007

Salaires	Effectifs n_i	Fréquences f_i %	Amplitudes a_i
[6000 – 7000[	12	24%	1000
[7000 – 8000[	10	20%	1000
[8000 – 9000[	15	30%	1000
[9000 – 10 000[	8	16%	1000
[10 000 – 11 000[	5	10%	1000
Total	50	100%	-

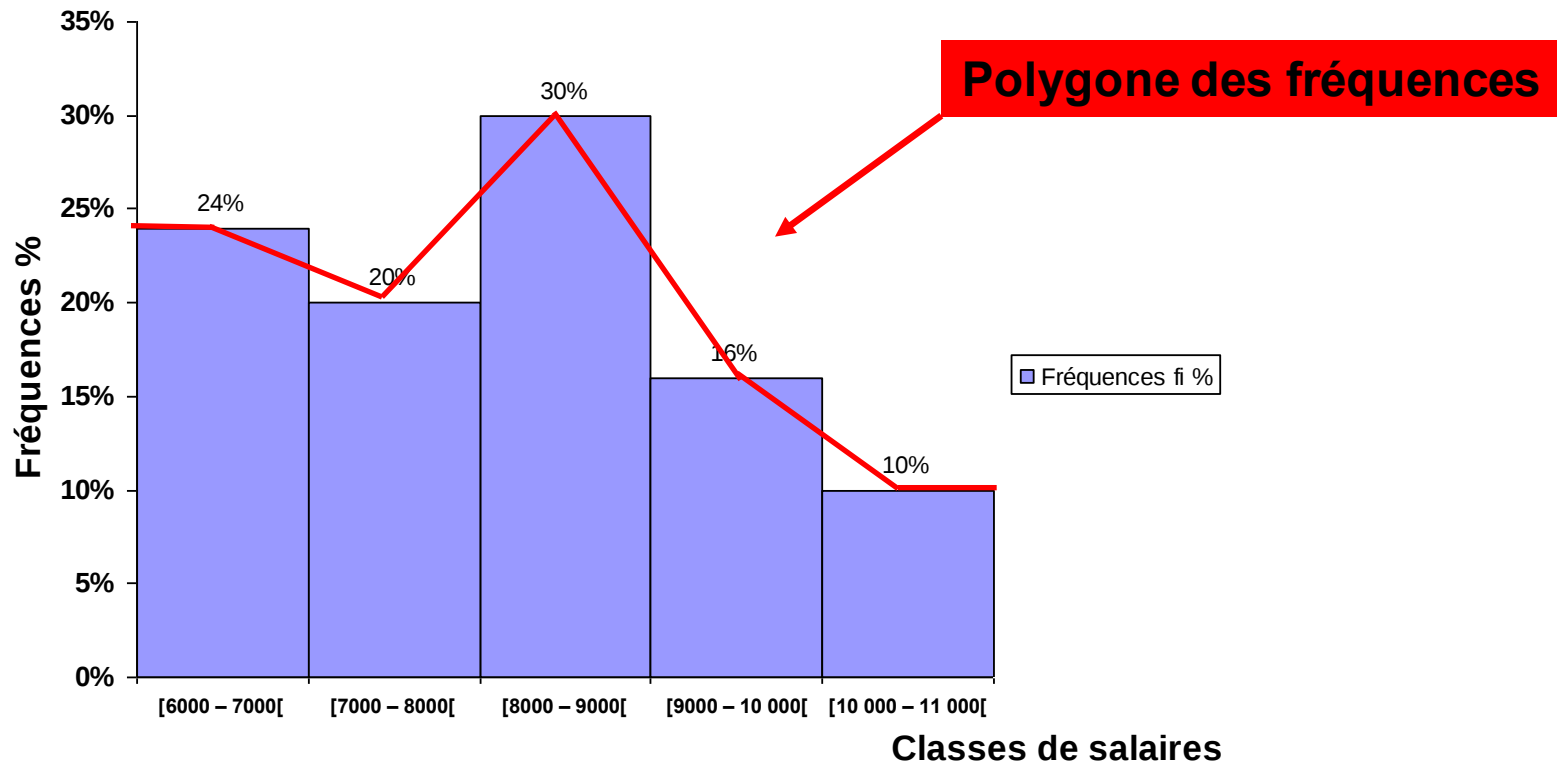


Les classes sont d'amplitudes égales

Chapitre III – Les représentations graphiques (19)

Représentations des caractères quantitatifs

Représentation graphique des salaires des 50 employés de l'entreprise « X »



Chapitre III – Les représentations graphiques (20)

Représentations des caractères quantitatifs

- *On obtient le polygone des fréquences en joignant les milieux des segments supérieurs de chaque rectangle de l'histogramme*
- *La propriété fondamentale du polygone des fréquences est qu'il conserve l'aire ou la surface de l'histogramme*
- *L'aire comprise entre le polygone des fréquences et l'axe des abscisses est la même que l'aire comprise dans l'histogramme*

Chapitre III – Les représentations graphiques (20)

Représentations des caractères quantitatifs

⇒ Cas de classes d'amplitudes inégales

- L'histogramme ne peut plus être construit exactement de la même manière
 - Les fréquences (effectifs) se rapportant à des classes d'amplitudes inégales ne sont plus comparables
 - Il faut dans ce cas effectuer une correction pour tenir compte des différences d'amplitude
 - Généralement, la correction se fait en calculant les fréquences (ou effectifs) par unité d'amplitude

$$\text{Fréquence corrigée par unité d'amplitude} = \frac{f_i}{a_i}$$

Chapitre III – Les représentations graphiques (21)

Représentations des caractères quantitatifs

- Sur l'axe des abscisses, sont portées ***les limites des classes***
- Sur l'axe des ordonnées, sont portées ***les fréquences*** (ou effectifs) ***corrigées*** correspondant à chaque classe
- ***La base de chaque rectangle de l'histogramme représente l'amplitude de classe***
- ***La hauteur de chaque rectangle de l'histogramme est proportionnelle à la fréquence (ou effectif) corrigée***

$$h_i = \frac{f_i}{a_i}$$

Chapitre III – Les représentations graphiques (22)

Représentations des caractères quantitatifs

- **Exemple** : Salaires des 50 employés de l'entreprise « X » en DH au 31/11/2007

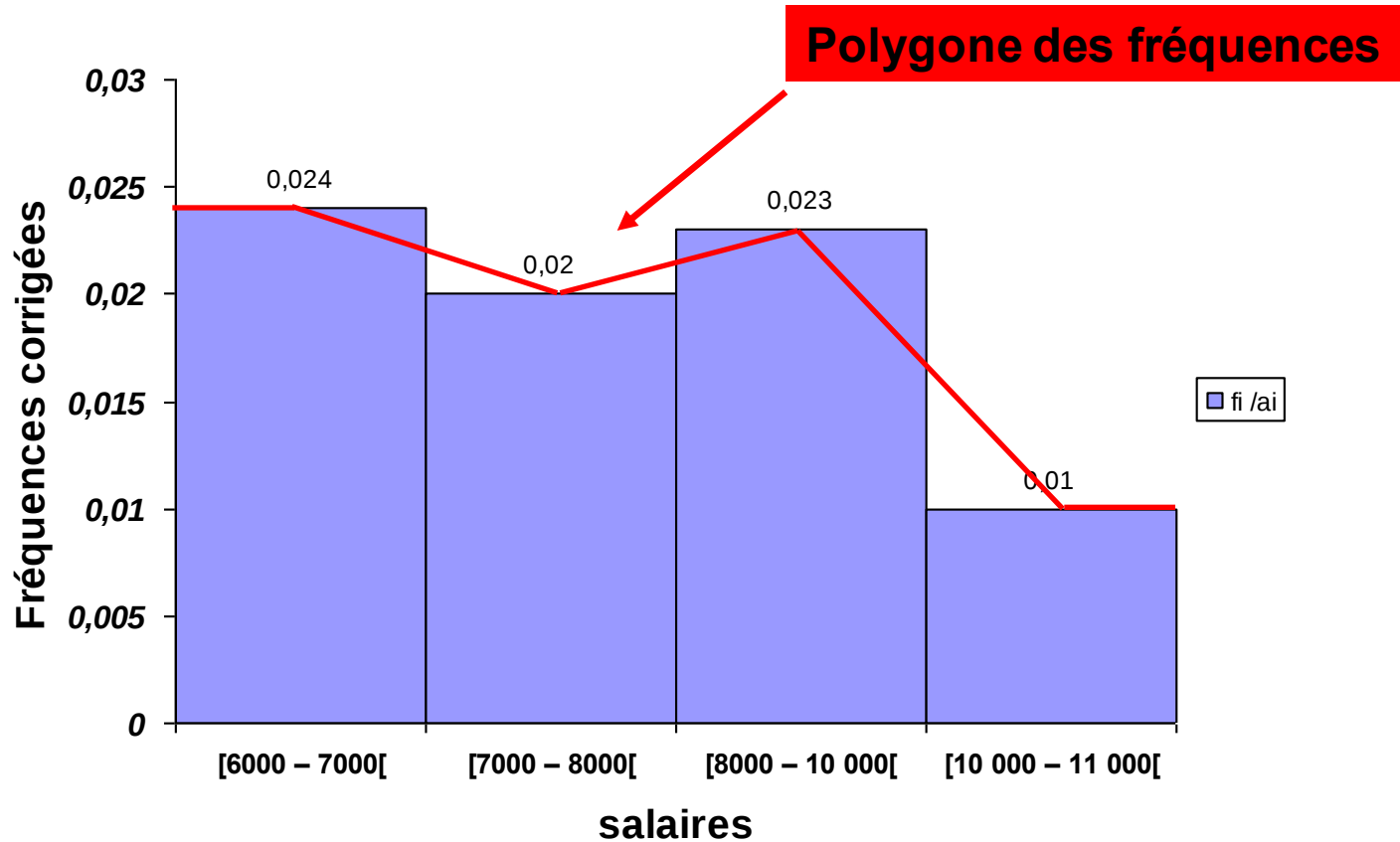
Salaires	n_i	$f_i\%$	a_i	f_i/a_i
[6000 – 7000[	12	24%	1000	0,024
[7000 – 8000[	10	20%	1000	0,02
[8000 – 10 000[	23	46%	2000	0,023
[10 000 – 11 000[	5	10%	1000	0,01
Total	50	100%	-	-

Les classes sont d'amplitudes inégales

Chapitre III – Les représentations graphiques (23)

Représentations des caractères quantitatifs

Représentation graphiques des salaires des 50 employés de l'entreprise « X »



Chapitre III – Les représentations graphiques (24)

Représentations des caractères quantitatifs

b) Représentation d'une distribution de fréquences (ou effectifs) cumulées

- La *fonction de répartition* $F(x)$ (ou $N(x)$) dans le cas de *caractère quantitatif continu* est représentée par *la courbe cumulative des fréquences* (effectifs)
- Les *fonctions de répartition* des *caractères quantitatifs continus* possèdent toutes les propriétés des fonctions de répartition des caractères discrets, excepté « *la continuité* »
- ➔ Les *fonctions de répartition* des caractères quantitatifs continus *sont continues à gauche et à droite*
 - ⇒ Dans chaque classe, on fait une *interpolation linéaire* : on relie les points extrêmes de chaque classe par un segment de droite
- ⇒ *La courbe cumulative est donc continue*

Chapitre III – Les représentations graphiques (25)

Représentations des caractères quantitatifs

■ **Exemple** : Salaires des 50 employés de l'entreprise « X » en DH au 31/11/2007

Salaires	n_i	$f_i\%$	F(x) « moins de »	F(x) « plus de »
[6000 – 7000[	12	24%	24%	100%
[7000 – 8000[	10	20%	44%	76%
[8000 – 9000[	15	30%	74%	56%
[9000 – 10 000[	8	16%	90%	26%
[10 000 – 11 000[	5	10%	100%	10%
Total	50	100%	-	-

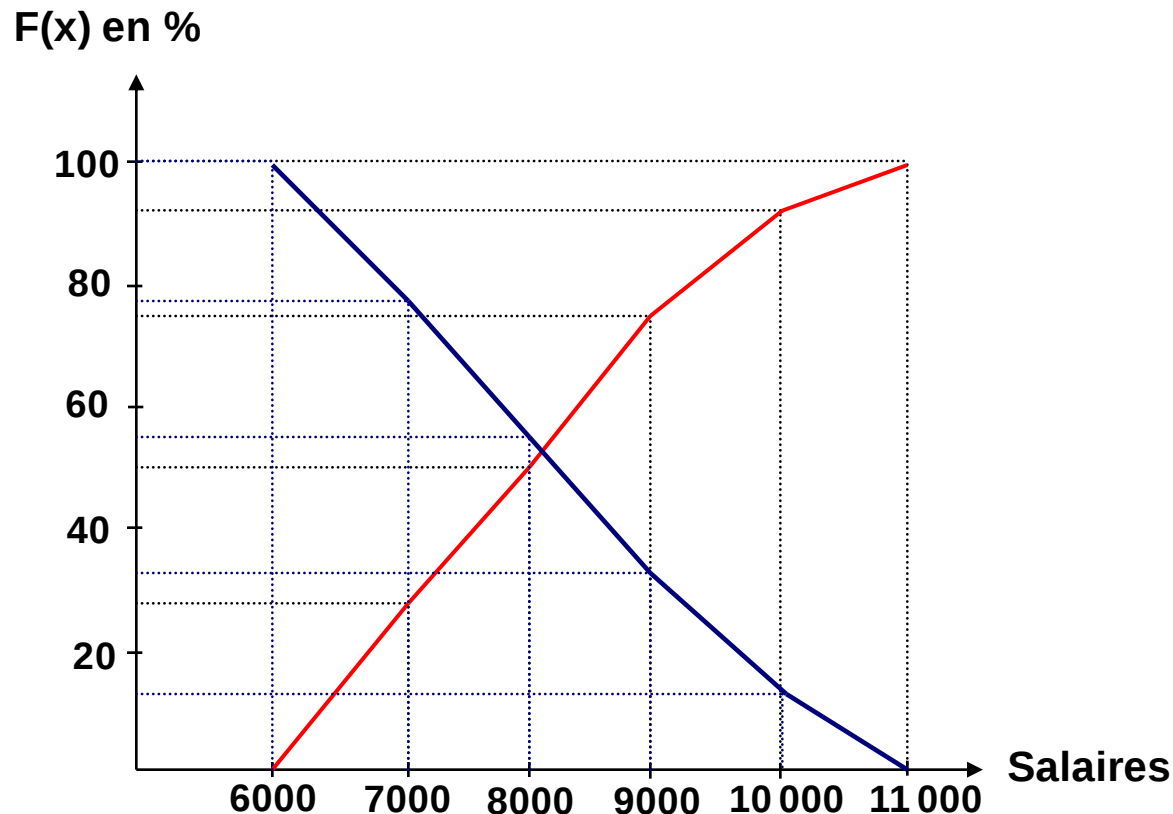
➔ Représenter graphiquement les fréquences cumulées

- ⇒ **F(x) « moins de »** : On prend pour abscisses les **limites supérieures des classes** et, pour ordonnées, les **fréquences cumulées** correspondantes
- ⇒ **F(x) « plus de »** : On prend pour abscisses les **limites inférieures des classes** et, pour ordonnées, les **fréquences cumulées** correspondantes

Chapitre III – Les représentations graphiques (26)

Représentations des caractères quantitatifs

Représentation graphique des courbes cumulatives des fréquences



Chapitre III – Les représentations graphiques (27)

Représentations des caractères à deux dimensions

- Les distributions statistiques à deux dimensions peuvent être représentées de différentes manières
 - ➔ Diagramme en tuyaux d'orgue
 - ➔ Diagramme circulaire
 - ➔ Stéréogramme
 - ➔ Etc.

1. Cas de caractère qualitatif

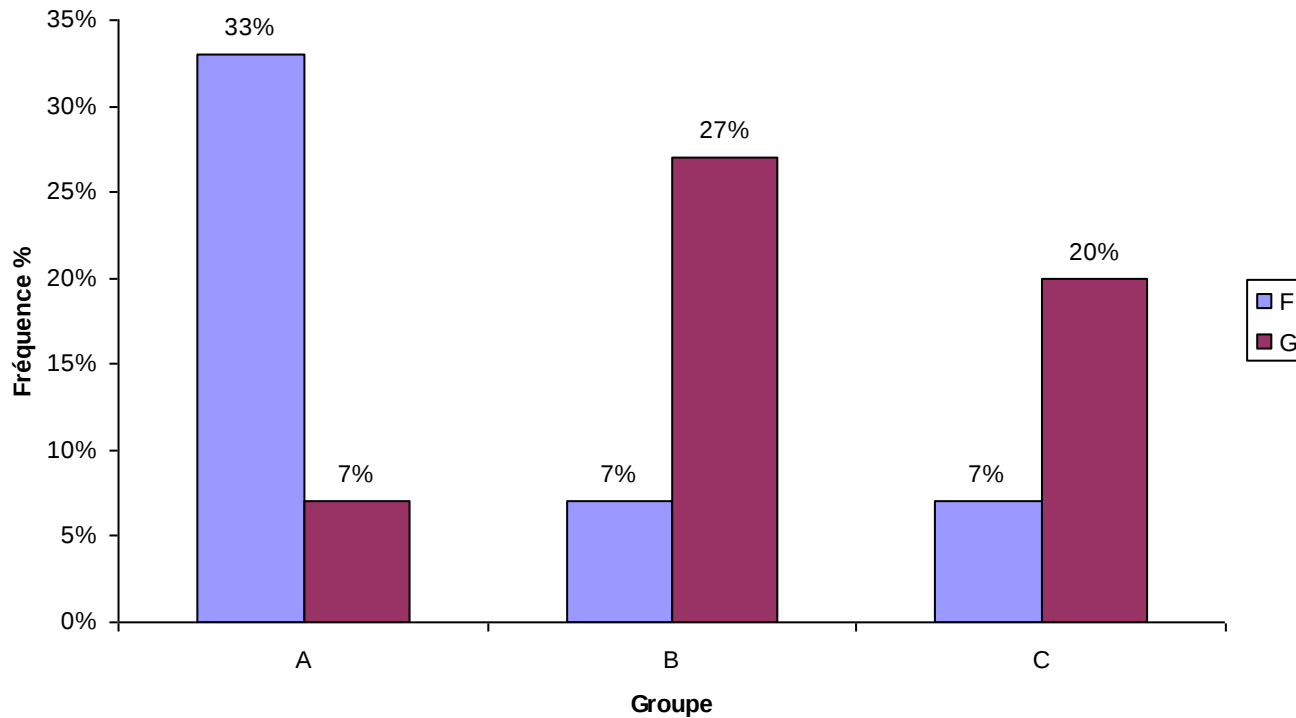
- ➔ Exemple : Répartition des élèves d'une classe selon le sexe et le groupe

Sexe \ Groupe				
	A	B	C	Total
F	5 (33%)	1 (7%)	1 (7%)	7 (47%)
G	1 (7%)	4 (27%)	3 (20%)	8 (53%)
Total	6 (40%)	5 (33%)	4 (27%)	15 (100%)

Chapitre III – Les représentations graphiques (28)

Représentations des caractères à deux dimensions

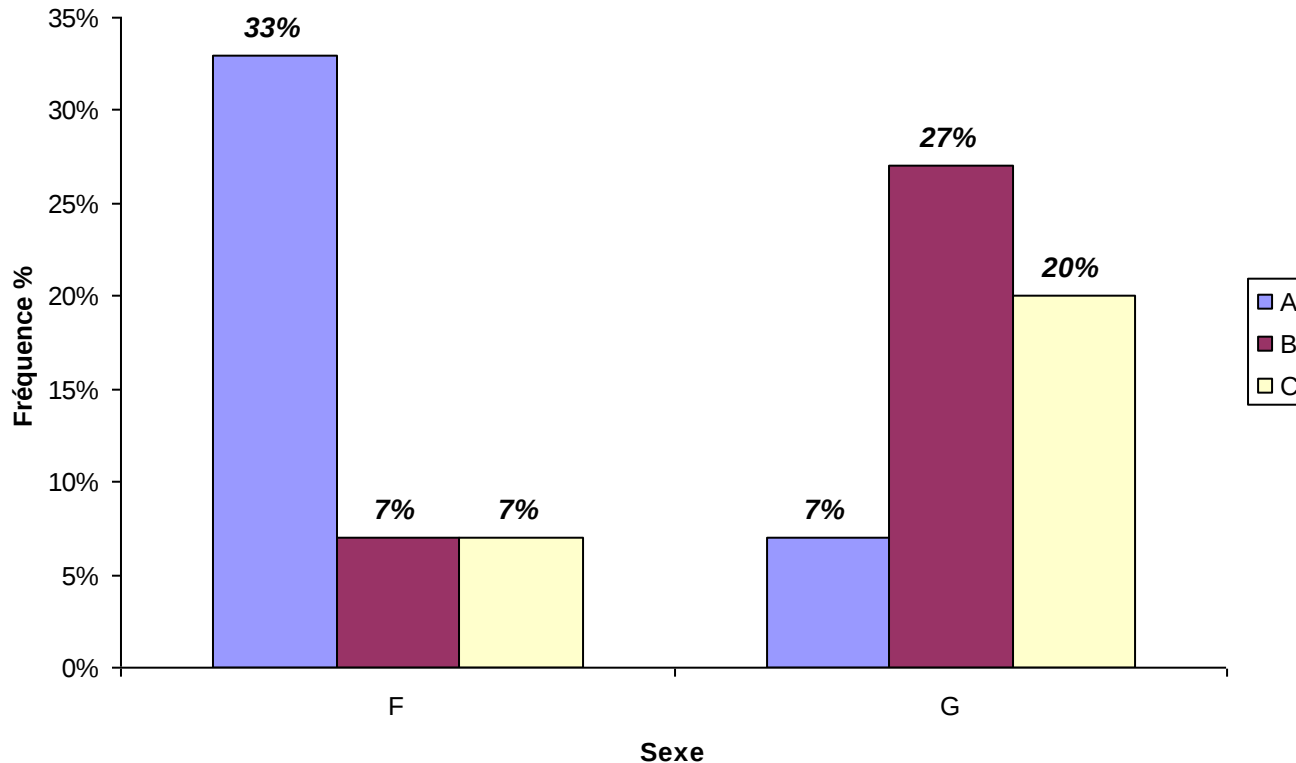
Répartition des élèves selon le sexe et le groupe



Chapitre III – Les représentations graphiques (29)

Représentations des caractères à deux dimensions

Répartition des élèves selon le sexe et le groupe



Chapitre III – Les représentations graphiques (30)

Représentations des caractères à deux dimensions

■ Représentation graphique des distributions conditionnelles

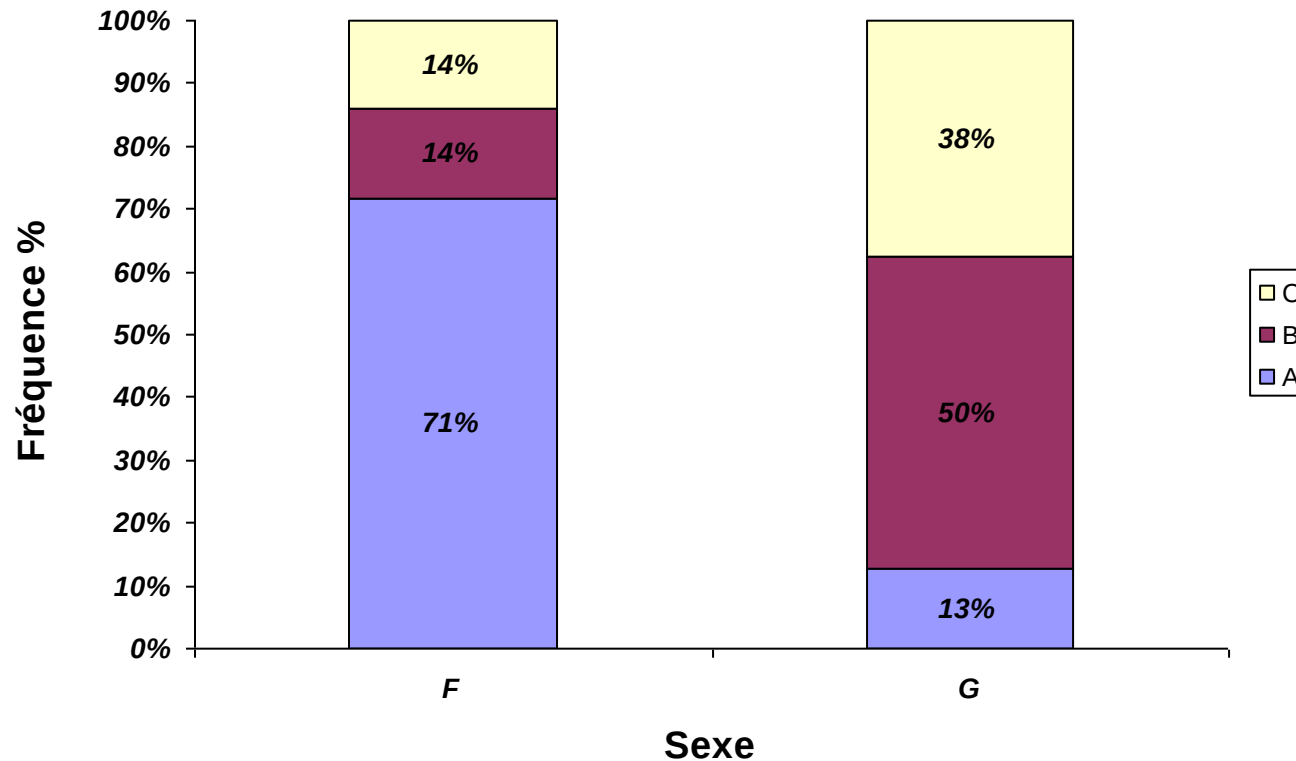
→ **Exemple 1** : Distributions de la variable GROUPE conditionnellement à la variable SEXE

SEXE \ GROUPE	GROUPE			
	A	B	C	Total
F	71% (5/7)	14% (1/7)	14% (1/7)	100%
G	13% (1/8)	50% (4/8)	38% (3/8)	100%
Total	40%	33%	27%	100%

Chapitre III – Les représentations graphiques (31)

Représentations des caractères à deux dimensions

Diagramme des fréquences de la variable GROUPE conditionnellement à la variable SEXE



Chapitre III – Les représentations graphiques (32)

Représentations des caractères à deux dimensions

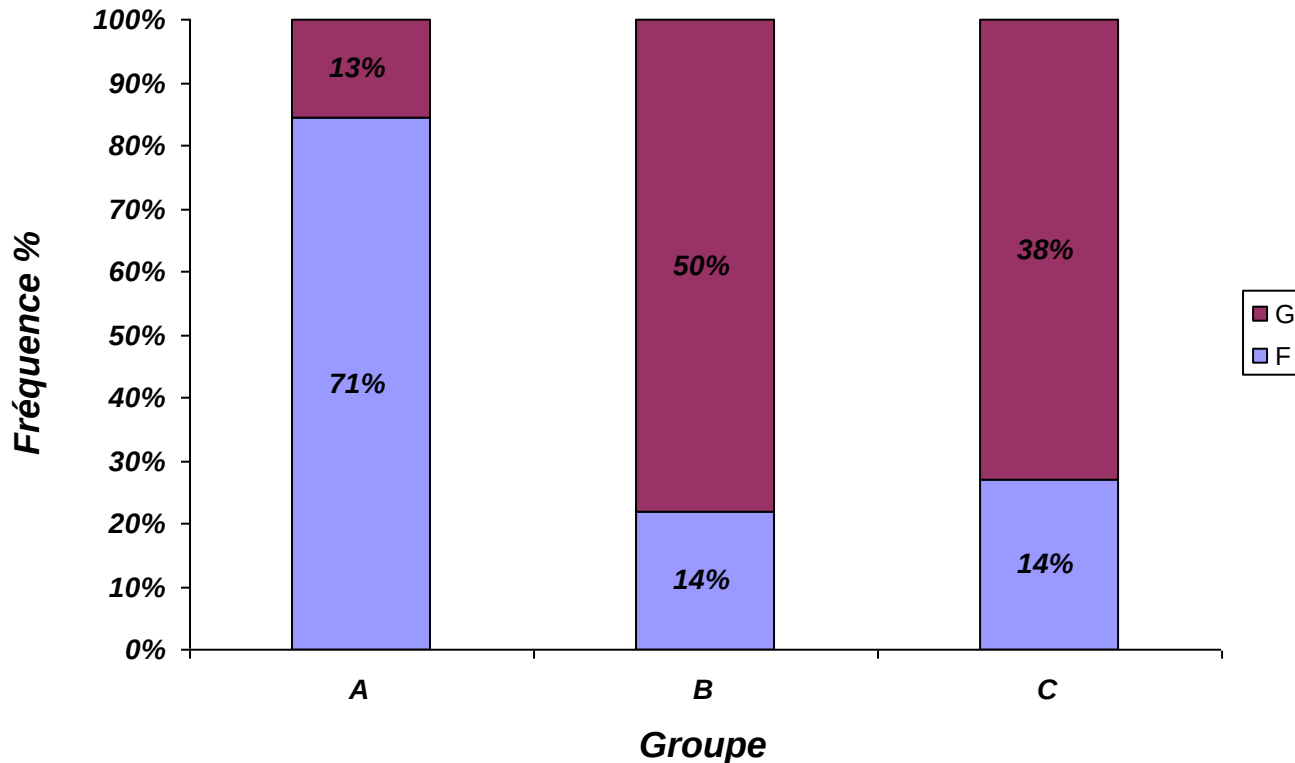
→ **Exemple 2** : Distributions de la variable SEXE conditionnellement à la variable GROUPE

SEXE \ GROUPE	GROUPE			
	A	B	C	Total
F	83% (5/6)	20% (1/5)	25% (1/4)	47%
G	17% (1/6)	80% (4/5)	75% (3/4)	53%
Total	100%	100%	100%	100%

Chapitre III – Les représentations graphiques (33)

Représentations des caractères à deux dimensions

Diagramme des fréquences de la variable SEXE conditionnellement à la variable GROUPE



Chapitre III – Les représentations graphiques (34)

Représentations des caractères à deux dimensions

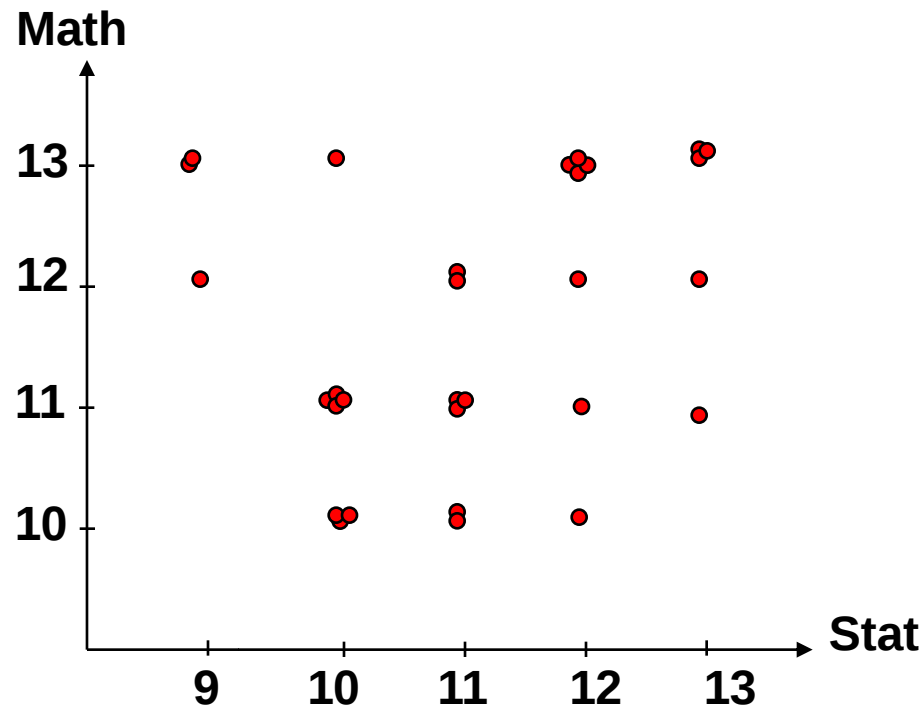
2. Cas de caractère quantitatif

- Les distributions statistiques à deux dimensions peuvent être représentées graphiquement sous forme de ***nuage de points*** dans un plan
 - Les points sont obtenus en représentant chaque couple d'observation $(x_i; y_i)$ par un point dans le plan
- **Exemple 1** : Distributions d'un groupe d'étudiants selon les notes de statistique et de mathématiques

Stat \ Math					
	10	11	12	13	Total
09	0	0	1	2	3
10	3	4	0	1	8
11	2	3	2	0	7
12	1	1	1	4	7
13	0	1	1	3	5
Total	6	9	5	10	30

Chapitre III – Les représentations graphiques (35)

Représentations des caractères à deux dimensions



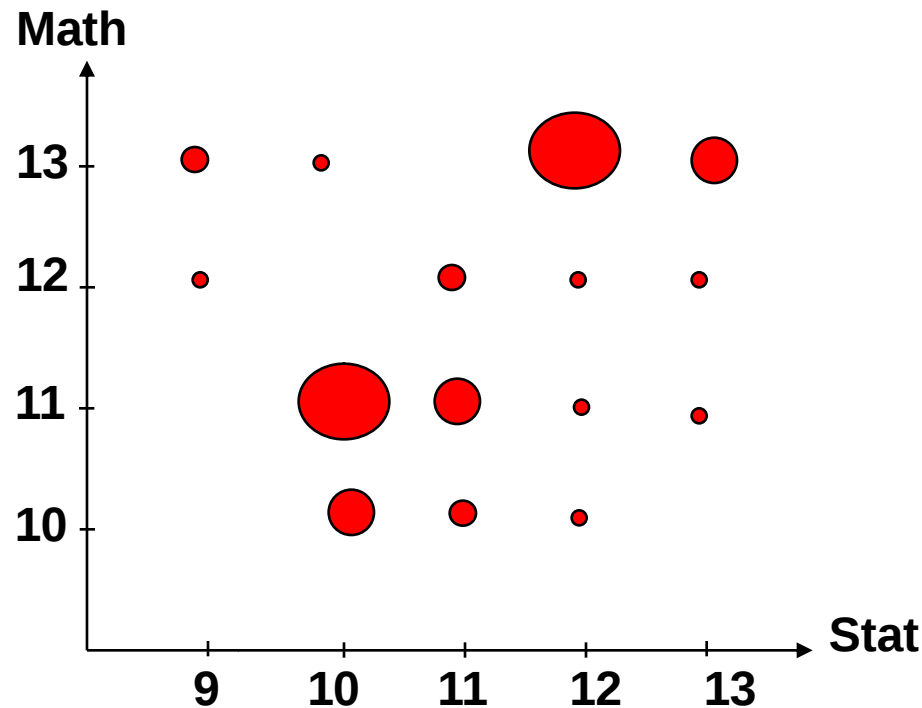
Représentation par nuage de points des étudiants selon leurs notes de statistique et de mathématiques

Chapitre III – Les représentations graphiques (36)

Représentations des caractères à deux dimensions

- On peut remplacer chaque point par un **cercle** délimitant une aire proportionnelle à l'effectif ou à la fréquence

Représentation des étudiants selon leurs notes de stat et de math



Chapitre III – Les représentations graphiques (37)

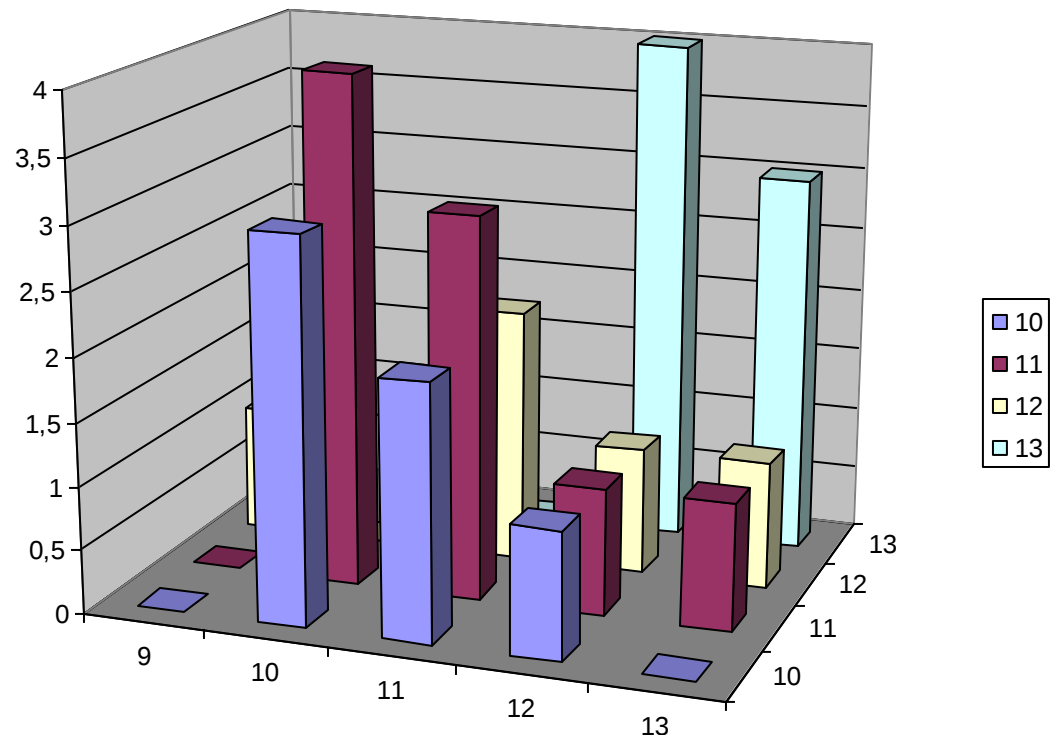
Représentations des caractères à deux dimensions

■ Autres représentations graphiques

➔ Stéréogramme

⇒ Le stéréogramme permet de faire des représentations 3D graphiques dans R^3

Représentation des étudiants
selon leurs notes de stat et de
math

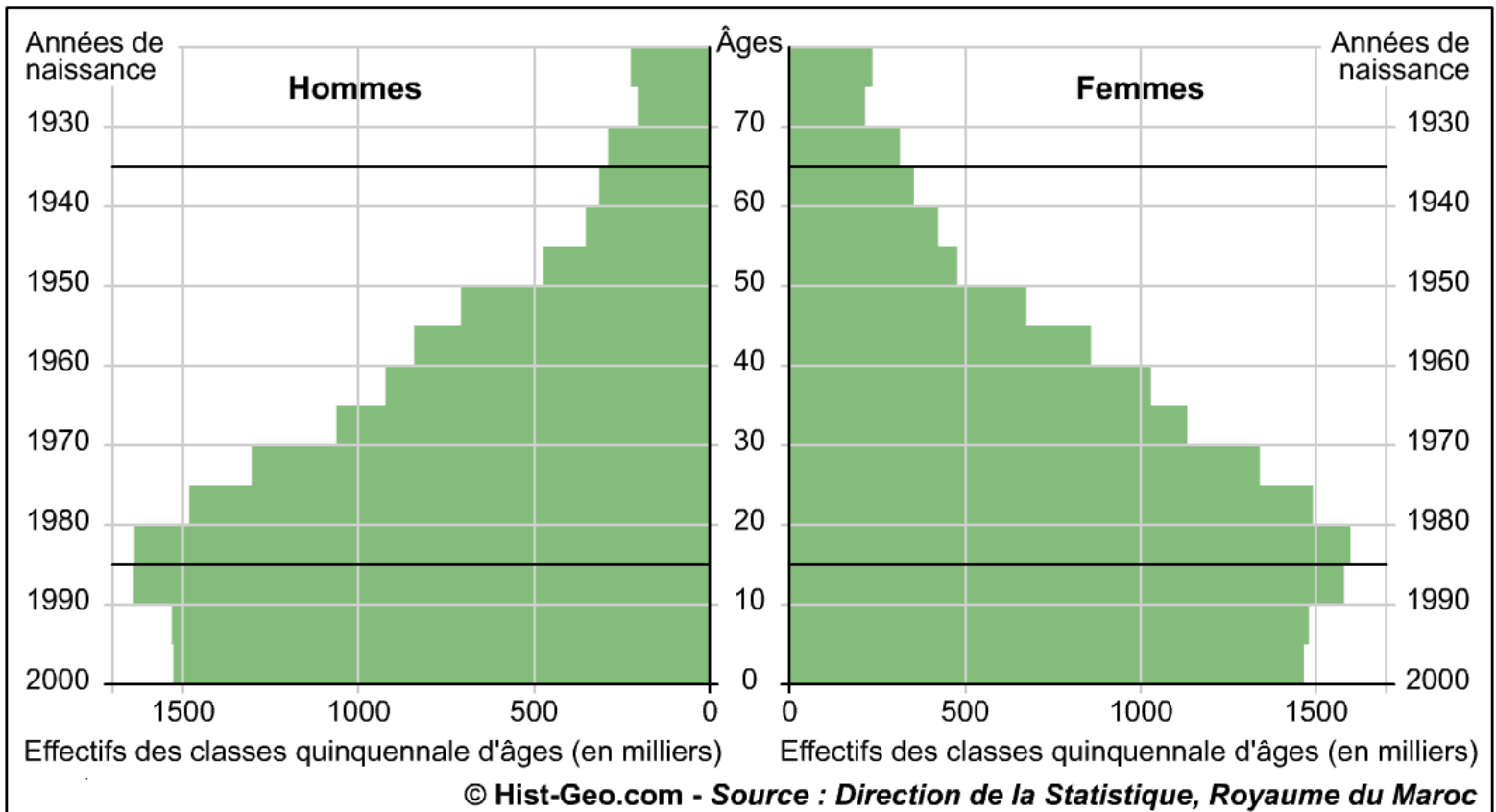


Chapitre III – Les représentations graphiques (38)

Représentations des caractères à deux dimensions

→ Pyramide des âges

Population du Maroc au 1er juillet 2000
(Répartition par âge et par sexe - Projections)

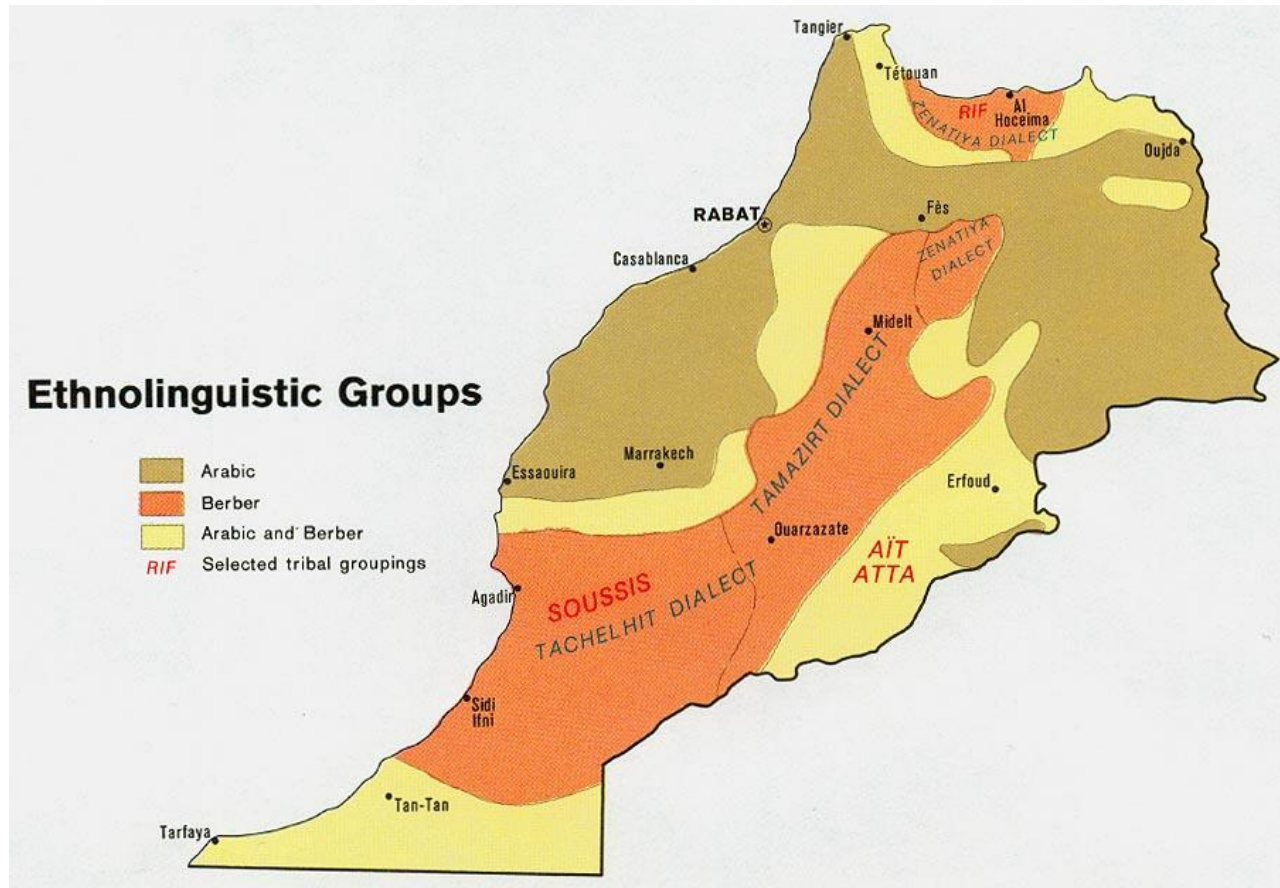


Chapitre III – Les représentations graphiques (39)

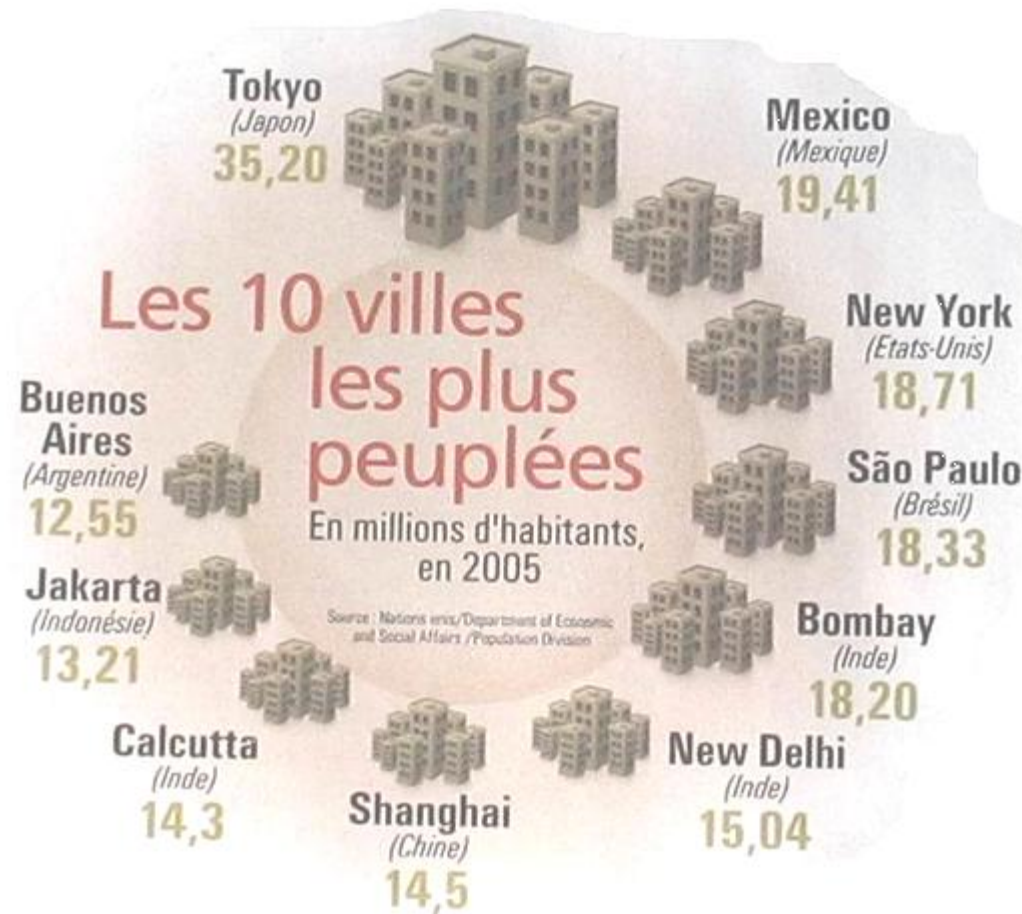
Conclusion

- Le chapitre 3 a porté sur les représentations graphiques des données statistiques
- Nous n'avons présenté que les représentations graphiques les plus courantes
- Au-delà de cette présentation non exhaustive, il existe des représentations appelées **cartogrammes** qui consistent à utiliser des cartes géographiques pour exprimer des distributions d'individus dans l'espace
- Il existe également des graphiques **figuratifs** où les phénomènes sont représentés par des objets en rapport avec le caractère étudié (Voiture pour la production de voitures; des sacs pour la production de blé, etc.)

Carte linguistique du maroc : arabe & berbère



Le diagramme figuratif



Chapitre III – Les séries chronologiques (1)

Introduction

